FXXMXCSGJLLYYY

王新洲 著

非线性模型参数估计理论与应用



全国优秀出版社
武汉大学出版社

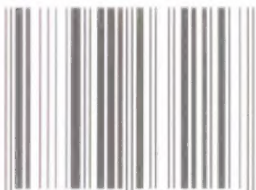


▪责任编辑／夏炽元
▪责任校对／张　昕
▪版式设计／支　笛
▪封面设计／何　琪



王新洲，工学博士，教授，博士生导师。1954年2月7日出生于湖北黄陂。1978年2月考入武汉测绘学院，就读于工程测量专业。1982年1月毕业分配到湖北黄石袁仓煤矿从事矿山测量工作。1983年9月调武汉电力学校任教。1991年2月在武汉测绘科技大学大地测量系获硕士学位，于同年9月继续攻读博士学位。1994年6月获工学博士学位留校后，于本年12月被评为副教授。1995年12月破格晋升为教授。1996年7月任武汉测绘科技大学地学测量工程学院副院长。1997年被评为国家测绘局跨世纪学术和技术带头人。1998年被遴选为博士生导师。1999年10月任武汉测绘科技大学科学技术处处长。2000年9月任武汉大学科学技术部副部长。1994年至1997年作为课题组主要骨干参加完成一项国家自然科学基金项目；1995年至2001年主持完成两项国家自然科学基金项目、一项国家测绘局八五科技攻关项目和一项国家测绘局测绘科技发展基金项目、一项国土资源部自由探索项目和一项国家重点实验室项目。2001年开始主持湖北省科技厅面上科研项目、湖北省科技厅重大科技攻关项目和国家高技术研究发展计划（863计划）项目，以及武汉市轨道交通一号线一期工程测绘保障体系的研究与建立等多项横向项目。在核心刊物及国际会议论文集上发表学术论文60余篇。出版学术专著一本、主编和参编全国高等学校试用教材4本、科技图书2本。1996年获国家测绘局测绘科技进步奖一等奖、1997年获国家测绘局第三届测绘教学奖一等奖、1998年获湖北省科学技术进步奖二等奖、1999年获中华人民共和国科学技术进步奖三等奖。

历任中国测绘学会工程测量分会副主任委员；中国测绘学会测绘教育委员会委员；全国高等学校测绘教育指导委员会委员；中国建筑学会工程勘察学术委员会委员；武汉测绘科技大学学报编辑委员会委员。现任中国测绘学会科普专业委员会副主任委员；中国测绘学会海洋测绘专业委员会委员；中国建筑学会工程勘察分会第六届委员；湖北省测绘学会常务理事兼科普专业委员会主任委员；武汉大学学报（信息科学版）编辑委员会委员；测绘工程编辑委员会委员；测绘遥感信息工程国家重点实验室兼职教授；国家GPS工程中心兼职教授；湖北警官学院兼职教授；山东建筑工程学院兼职教授；重庆大学兼职教授。



ISBN 7-307-03515-4/O·262
定价：15.00元

非线性模型参数估计理论与应用

王新洲　著

■ 国家自然科学基金项目(49474204;49874002)资助
■ 国家高技术研究发展计划(863 计划)项目(2001AA135081)资助
■ 国家测绘局测绘科技发展基金项目(99010)资助
■ 测绘遥感信息工程国家重点实验室开放研究基金(测绘基础研究)项目((00)0204)资助

武汉大学出版社

图书在版编目(CIP)数据

非线性模型参数估计理论与应用/王新洲著.—武汉：武汉大学出版社，2002.6

ISBN 7-307-03515-4

Ⅰ.非…　Ⅱ.王…　Ⅲ.非线性—数学模型—参数估计　Ⅳ.O211.67

中国版本图书馆CIP数据核字(2002)第018702号

责任编辑：夏炽元　　责任校对：张　昕　　版式设计：支　笛

出版：武汉大学出版社　(430072　武昌　珞珈山)
(电子邮件：wdp4@whu.edu.cn　网址：www.wdp.whu.edu.cn)
发行：新华书店湖北发行所
印刷：湖北恒吉印务有限公司
开本：787×1092　1/8　印张：9.125　字数：215千字
版次：2002年6月第1版　　2002年6月第1次印刷
ISBN 7-307-03515-4/O·262　　定价：15.00元

序

我们知道,所谓参数估计,它是数理统计中的一个名词,其含义是根据含有误差的观测向量,依一定的数学模型和准则求解未知参数.实际上在测量工作中这就是平差问题。通常无论是在数理统计,还是测量平差中,依据线性数学模型的参数估计理论已是非常成熟而系统的理论,是其主要研究和应用的对象.可是一般情况下,包括我们测量工作,非线性数学模型的出现要比线性模型频繁得多,对此,传统的方法是将数学模型进行线性化近似,即将非线性问题转化成线性问题来处理.在过去测量精度不高的情况下,可以忽略由此线性化近似而引起的数学模型的误差.可是这种线性化近似已不能满足现代高精度测量工作的要求,它完全可能导致参数估计结果的精度扭曲.为此本书作者于1994~2001年期间在国家自然科学基金项目"非线性随机模型估计理论并用于全球定位系统"等项目的资助下,深入系统地研究了非线性模型的参数估计理论及其在测量数据处理中的应用,取得了一系列高水平的成果.本书就是在这些研究成果的基础上,通过整理扩充而成的.其主要特点,一是直接对非线性模型进行处理,从而精确地揭示了非线性模型中估计量的统计性质,从理论上解决了由于非线性模型的线性化近似所引起的各种问题;二是紧密结合测量数据处理实际,抓住非线性模型参数估计的算法和精度评定两个基本问题进行深入浅出的论述,在严谨的理论分析的基础上给出了便于实际应用的公式.所以本书既有较高的理论水平,又有较大的实用价值。

本书作者王新洲教授从事测量数据处理理论与应用的教学与科研工作十余年,具有较强的科研能力和较高的业务水平,在这一学科领域既取得了一些高水平成果,也积累了较丰富的实践经验,同时具有很强的写作能力.本书不仅文字流畅、层次分明,而且论证严谨、逻辑性强.我相信这会是一本深受读者欢迎的好书!

2002年5月5日

前　言

非线性模型参数估计理论是国家自然科学基金委员会于 1994 年在自然科学学科发展战略调整调查报告《大地测量学》一书中提出的大地测量学学科发展面临的重大基础理论问题之一，也是国际大地测量协会大地测量数学与物理基础“统计学”专题研究组 1991～1995 年的重点研究内容之一．因为测量上大量的数学模型都是非线性模型，很多观测方程都具有很强的非线性性，对级数展开的初值点十分敏感，所以需要研究模型空间的容许曲率问题，以及在非线性函数空间的平差理论和可普及的实用方法．

对于大地测量中大量的非线性模型，传统的方法是进行线性近似，即将其展开为泰勒级数，取至一次项，而略去二次以上各项．如此线性近似，必然会引起模型误差．由于过去测量精度不高，线性近似所引起的模型误差一般小于观测误差，故线性近似所引起的模型误差可忽略不计．随着科学技术的不断发展，现在的测量精度已大大提高，致使线性近似所引起的模型误差与观测误差相当，甚至还会大于观测误差．因此，用近似的理论、模型、方法去处理具有很高精度的观测结果，从而导致精度损失，显然是不合理的．现代科学技术要求估计结果的精度尽可能提高．这样，传统的线性近似的方法就不能满足当今科学技术的要求．另外，有些非线性模型对参数的近似值十分敏感，若近似值的精度较差，线性近似时就会产生较大的模型误差．此时用线性模型的精度评定理论去评定估计结果的精度，会得到一些虚假的优良统计性质，人为地拔高了估计结果的精度．

数理统计中对非线性参数估计的研究始于 20 世纪 60 年代初期．但在开始的 20 年中进展并不快．直到 20 世纪 80 年代初，加拿大统计学家 Bates 和 Watts 引进曲率度量后此理论才得到较快的发展．

本书的内容主要来自我们的研究成果．同时参考了国内外一系列资料和最新研究成果．本书的主要内容包括：线性模型参数估计理论的回顾；非线性模型非线性强度的度量理论；非线性最小二乘估计的各种算法及其效率比较；非线性最小二乘估计量的统计性质及精度评定；非线性模型的其他估计方法等．本书从理论上解决了线性近似所引起的问题，是将非线性科学引入测绘领域的一个开端．

本书附录 A 中介绍的立体阵，在本书中经常用到，建议读者事先熟练掌握它．

由于作者水平有限，谬误之处在所难免，恳请各位专家和读者不吝赐教．

作　者

2002 年 2 月 22 日于武昌

目　录

第一章 参数估计概述

参数估计(Parameter Estimation)是一种基本的统计推断形式,也是数理统计学的一个重要分支[1](成平,陈希孺等,1985),更是测量数据处理(Surveying Data Processing)理论的重要组成部分.由于迄今为止,参数估计的一系列成果主要集中在线性模型(Line Model),而且线性模型参数估计理论是非线性模型(Nonlinear Model)参数估计的基础,它们之间有很多联系.所以,本章首先简要回顾线性模型参数估计中的基本理论和方法.

第一节 线性模型参数估计理论的回顾

线性模型是数理统计学中发展比较早的分支之一.关于它的参数估计问题,可以追溯到18世纪初[2][3](王松桂,1987).1806年著名数学家A.M.Legendre在《决定彗星轨道新方法》中从代数观点提出了最小二乘法(Least Square).而早在1794年,年仅17岁的高斯(C.F.Gauss)就提出用最小二乘法从带有误差的观测值中找出待定量的最优值.但高斯只到1809年才在《天体运动的理论》中正式发表他的方法[4](武汉测绘科技大学测量平差教研室,1996).后来,马尔可夫(A.A.Markov)于1900年证明了最小二乘估计的方差(Variance)最小的性质.形成了著名的Gauss-Markov定理,从而奠定了最小二乘法在线性模型参数估计中的地位.

1944年,R.C.Bose引入了可估函数的概念,加之广义逆矩阵的应用,使得设计矩阵为列降秩的线性模型参数估计理论表述得更加严密而简洁[2][3](王松桂,1987).20世纪60年代中期开始研究观测误差的方差-协方差矩阵为奇异矩阵的线性模型参数估计问题.Goldman和Zelen率先提出了用满秩线性变换把估计模型化为方差-协方差矩阵为$\sigma^2 I$,且带有线性约束的情形.后来C.R.Rao采用推广最小二乘法的途径,提出了所谓的“最小二乘法统一理论”.该统一理论既适合于设计矩阵列满秩或列降秩,又适合于观测误差的方差-协方差矩阵奇异的情况.这些结果构成了线性模型最小二乘估计理论的基本内容[2][3](王松桂,1987).

一、线性模型参数的最小二乘估计

设线性模型为

$$\begin{cases} L = BX + \Delta \\ \mathrm{E}(\Delta) = 0 \\ \mathrm{Var}(\Delta) = \sigma^2 Q_{LL} \end{cases} \tag{1-1-1}$$

式中:L为$n \times n$的观测向量;B为$n \times t$的设计矩阵;X为$t \times 1$的未知参数向量;Δ为$n \times 1$的观测误差向量;σ^2为单位权方差因子;Q_{LL}为$n \times n$的协因数矩阵,且$Q_{LL} > 0$.

线性模型(1-1-1)式是测量平差(Surveying Adjustment)中最常用的数学模型.其中第一式通常称为观测方程(Observation Equation),测量平差基础中称之为函数模型(Function Model).而将(1-1-1)式中的第二、第三式称之为随机模型[4](Stochastic Model)(武汉测绘科技大学测量平差教研室,1996).

记未知参数 X 的最小二乘估计为$\hat{X}_{LS}$;残差(Residual)向量为 V.用 $\hat{X}_{LS}$和 V 代替X 和 Δ,则观测方程变为

$$V = B\hat{X}_{LS} - L \tag{1-1-2}$$

测量上通常称(1-1-2)式为误差方程(Error Equation).

再记观测值的权矩阵为

$$P = P' = Q_{LL}^{-1} \tag{1-1-3}$$

线性模型(1-1-1)式中未知参数 X 的最小二乘估计,就是寻求 X 的一个估值,使

$$V'PV = (B\hat{X}_{LS} - L)'P(B\hat{X}_{LS} - L) = \min \tag{1-1-4}$$

为此,将(1-1-4)式对 $\hat{X}_{LS}$求一阶导数,并令其为零,得

$$\frac{\mathrm{d}V'PV}{\mathrm{d}\hat{X}_{LS}} = V'PB = 0$$

即

$$B'PV = 0 \tag{1-1-5}$$

将(1-1-2)式代入(1-1-5)式,得

$$B'PB\hat{X}_{LS} - B'PL = 0 \tag{1-1-6}$$

测量平差中称(1-1-6)式为法方程(Normal Equation).

法方程在最小二乘估计中起着非常重要的作用.

当设计矩阵 B 列满秩时,由法方程(1-1-6)式可解得未知参数 X 的最小二乘解为

$$\hat{X}_{LS} = (B'PB)^{-1}B'PL \tag{1-1-7}$$

当设计矩阵 B 列降秩时,法方程(1-1-6)的系数矩阵 $B'PB$ 也降秩.在这种情况下,法方程(1-1-6)式的惟一解为

$$\hat{X}_{LS} = (B'PB)^{+}B'PL \tag{1-1-8}$$

测量平差中称这种情况为秩亏自由网平差(Adjustment of Free-Network with Rank Deficiency).

二、线性模型参数最小二乘估计的统计性质

1.无偏性

对(1-1-7)式所确定的最小二乘估计取数学期望,得

$$\mathrm{E}(\hat{X}_{LS}) = (B'PB)^{-1}B'P\mathrm{E}(L) \tag{1-1-9}$$

由(1-1-1)式可得

$$\mathrm{E}(L) = BX + \mathrm{E}(\Delta) = BX \tag{1-1-10}$$

将(1-1-10)式代入(1-1-9)式,得

$$\mathrm{E}(\hat{X}_{LS}) = (B'PB)^{-1}B'PBX = X \tag{1-1-11}$$

(1-1-11)式表明最小二乘估计 $\hat{X}_{LS}$是未知参数 X 的无偏估计.

2.方差最小性

设线性模型(1-1-1)式中参数 X 的最优线性无偏估计(BLUE)为

$$\hat{X} = ML \tag{1-1-12}$$

式中：M 为 $t \times n$ 的待定系数矩阵.

对(1-1-12)式取数学期望，得

$$\mathrm{E}(\hat{X}) = M\mathrm{E}(L) = MBX$$

因为 $\hat{X} = ML$ 为最优线性无偏估计，所以有

$$MB = I \tag{1-1-13}$$

式中：I 为 $t \times t$ 的单位阵.

由(1-1-12)式应用协因数传播律，得

$$MQ_{LL}M' = \min \tag{1-1-14}$$

下面来求满足条件(1-1-13)式，又使(1-1-14)式成立的矩阵 M. 为此，组成函数

$$\Phi = MQ_{LL}M' + 2(MB - I)K \tag{1-1-15}$$

(1-1-15)式对 M 求偏导数，并令其为零，得

$$\frac{\partial \Phi}{\partial M} = 2Q_{LL}M' + 2BK = 0$$

即

$$MQ_{LL} + K'B' = 0$$

于是有

$$M = -K'B'Q_{LL} = -K'B'P \tag{1-1-16}$$

将(1-1-16)式代入(1-1-13)式，得

$$-K'B'PB = I$$

所以有

$$K = -(B'PB)^{-1} \tag{1-1-17}$$

将(1-1-17)式代入(1-1-16)式得最优线性无偏估计的系数矩阵为

$$M = (B'PB)^{-1}BP \tag{1-1-18}$$

将(1-1-18)式代入(1-1-12)式，得线性模型中未知参数 X 的最优线性无偏估计为

$$\hat{X} = (B'PB)^{-1}B'PL = \hat{X}_{\mathrm{LS}} \tag{1-1-19}$$

所以 $\hat{X}_{\mathrm{LS}}$为线性模型参数的最优线性无偏估计. 由此证明线性模型最小二乘估计 $\hat{X}_{\mathrm{LS}}$ 具有最小方差.

三、单位权方差因子 σ^2 的惟一非负最优二次无偏估计

线性模型(1-1-1)式中单位权方差因子 σ^2 由下式估计：

$$\hat{\sigma}^2 = \frac{V'PV}{n-t} \tag{1-1-20}$$

式中：n 为观测值的个数；t 为必要观测的个数.

可以证明，$\hat{\sigma}^2$ 是单位权方差因子 σ^2 的惟一最小方差非负二次无偏估计. 下面仅证 $\hat{\sigma}^2$ 的无偏性. 至于 $\hat{\sigma}^2$ 的其他性质的证明，有兴趣的读者可参阅《线性牛型的理论及其应用》[2][3](王松桂，1987).

为了证明 $\hat{\sigma}^2$ 的无偏性，我们先证明一个定理.

定理 1-1-1 设 y 为 $n \times 1$ 的随机向量，$\mathrm{E}(y) = \mu$，$\mathrm{Var}(y) = D$，M 为 $n \times n$ 的对称方

阵,则

$$E(Y'MY)=\mathrm{tr}(MD)+\mu'M\mu \tag{1-1-21}$$

式中:tr(Z)表示求方阵 Z 的迹.

证明: $E(Y'MY)$

$$\begin{aligned}
&=E[(y-\mu)'M(y-\mu)+\mu'My+y'M\mu-\mu'M\mu]\\
&=E[(y-\mu)'M(y-\mu)]+2E(y'M\mu)-\mu'M\mu\\
&=E\{\mathrm{tr}[(y-\mu)'M(y-\mu)]\}+\mu'M\mu\\
&=E\{\mathrm{tr}[M(y-\mu)(y-\mu)']\}+\mu'M\mu\\
&=\mathrm{tr}\{ME[(y-\mu)(y-\mu)']\}+\mu'M\mu\\
&=\mathrm{tr}(MD)+\mu'M\mu
\end{aligned}$$

定理 1-1-1 得证.

下面再来证明 $\hat{\sigma}^2$ 的无偏性.

因为由(1-1-2)式,顾及(1-1-1)式,有

$$\begin{aligned}
E(V)&=B(\hat{X}_{LS})-E(L)\\
&=BX-BX=0
\end{aligned} \tag{1-1-22}$$

$$\mathrm{Var}(V)=\sigma^2(Q_{LL}-B(B'PB)^{-1}B') \tag{1-1-23}$$

于是,由定理 1-1-1,顾及(1-1-3)式,有

$$\begin{aligned}
E(V'PBV)&=\sigma^2\mathrm{tr}\{P(Q_{LL}-B(B'PB)^{-1}B')\}\\
&=\sigma^2[\mathrm{tr}(PQ_{LL})-\mathrm{tr}(PB(B'PB)^{-1}B')]\\
&=\sigma^2[\mathrm{tr}(\underset{n\times n}{I})-\mathrm{tr}(\underset{t\times t}{I})]\\
&=\sigma^2(n-t)
\end{aligned}$$

所以

$$E(\hat{\sigma}^2)=\frac{E(V'PV)}{n-t}=\sigma^2 \tag{1-1-24}$$

因为 $\hat{\sigma}^2$ 是基于 X 的最小二乘估计 $\hat{X}_{LS}$ 导出的,所以有文献称 $\hat{\sigma}^2$ 为 σ^2 的最小二乘估计.

四、线性模型最小二乘估计的序贯解法

当我们已经得到线性模型的最小二乘估计后,又因某种原因增加了一个或多个观测值时,为了利用以前的估计结果,不再额外增加估计工作量,就可以采用最小二乘估计的序贯解法.为此,将线性模型(1-1-1)式中的观测向量 L 分为两组,记为 L_{k-1} 和 L_k.它们的权矩阵分别记为 P_{k-1} 和 P_k,并设 L_{k-1} 与 L_k 不相关.则误差方程(1-1-2)式可写为

$$\begin{cases}
V_{k-1}=B_{k-1}\hat{X}_{LS}-L_{k-1}\\
V_k=B_k\hat{X}_{LS}-L_k
\end{cases} \tag{1-1-25}$$

式中:L_{k-1} 为 $n_{k-1}\times 1$ 的观测向量;L_k 为 $n_k\times 1$ 的观测向量;V_{k-1} 为 $n_{k-1}\times 1$ 的残差向量;V_k 为 $n_k\times 1$ 的残差向量;B_{k-1} 为 $n_{k-1}\times 1$ 的设计矩阵,$n_{k-1}>t$,且有 $\mathrm{rk}(B_{k-1})=t$(注:rk(B)表示矩阵 B 的秩),即 B_{k-1} 列满秩;B_k 为 $n_k\times 1$ 的设计矩阵.与(1-1-25)式相应

的权矩阵为 $P=\begin{pmatrix} P_{k-1} & 0 \\ 0 & P_k \end{pmatrix}$.

单独用(1-1-25)式中的第一式估计 $\hat{X}_{LS}$,得

$$\hat{X}_{LS}^{k-1}=(B_{k-1}'P_{k-1}B_{k-1})^{-1}B_{k-1}'P_{k-1}L_{k-1} \tag{1-1-26}$$

而用(1-1-25)式整体估计 $\hat{X}_{LS}$,得

$$\begin{aligned}\hat{X}_{LS}&=(B_{k-1}'P_{k-1}B_{k-1}+B_k'P_kB_k)^{-1}(B_{k-1}'P_{k-1}L_{k-1}+B_k'P_kL_k)\\&=\hat{X}_{LS}^{k-1}+(B_{k-1}'P_{k-1}B_{k-1}+B_k'P_kB_k)^{-1}B_k'P_k(L_k-B_k\hat{X}_{LS}^{k-1})\end{aligned} \tag{1-1-27}$$

令

$$\begin{cases} Q_{X_{k-1}}=(B_{k-1}'P_{k-1}B_{k-1})^{-1} \\ Q_X=(Q_{X_{k-1}}^{-1}+B_k'P_kB_k)^{-1} \end{cases} \tag{1-1-28}$$

则(1-1-27)式可写为

$$\begin{aligned}\hat{X}_{LS}&=\hat{X}_{LS}^{k-1}+Q_XB_k'P_k(L_k-B_k\hat{X}_{LS}^{k-1})\\&=\hat{X}_{LS}^{k-1}+J(L_k-B_k\hat{X}_{LS}^{k-1})\end{aligned} \tag{1-1-29}$$

式中:

$$J=Q_XB_k'P_k \tag{1-1-30}$$

由(1-1-28)式的第二式可得

$$Q_{X_{k-1}}^{-1}=Q_X^{-1}-B_k'P_kB_k \tag{1-1-31}$$

将(1-1-31)式两边左乘 Q_X,得

$$Q_XQ_{X_{k-1}}^{-1}=I-JB_k \tag{1-1-32}$$

将(1-1-32)式两边右乘 $Q_{X_{k-1}}$,得

$$Q_X=Q_{X_{k-1}}-JB_kQ_{X_{k-1}} \tag{1-1-33}$$

由(1-1-28)式的第二式,应用矩阵反演公式,可得

$$Q_X=Q_{X_{k-1}}-Q_{X_{k-1}}B_k'(P_k^{-1}+B_kQ_{X_{k-1}}B_k')^{-1}B_kQ_{X_{k-1}} \tag{1-1-34}$$

比较(1-1-33)式与(1-1-34)式,得

$$J=Q_XB_k'P_k=Q_{X_{k-1}}B_k'(P_k^{-1}+B_kQ_{X_{k-1}}B_k')^{-1} \tag{1-1-35}$$

将(1-1-35)式代入(1-1-29)式,得最小二乘序贯解的递推公式为

$$\hat{X}_{LS}=\hat{X}_{LS}^{k-1}+Q_{X_{k-1}}B_k'(P_k^{-1}+B_kQ_{X_{k-1}}B_k')^{-1}(L_k-B_k\hat{X}_{LS}^{k-1}) \tag{1-1-36}$$

当 $n_k=1$ 时,$(P_k^{-1}+B_kQ_{X_{k-1}}B_k')$ 和 $L_k-B_k\hat{X}_{LS}^{k-1}$ 都是一个数. 因此在这种情况下,最小二乘序贯解计算非常简单,所以总是假定 $n_k=1$,逐次递推,直到求出整体估计时的最后结果[4](武汉测绘科技大学测量平差教研室,1996). 最小二乘序贯解法可应用于测量控制网平差后又因某种原因增加观测值的情况.

五、带有线性约束的线性模型参数估计

以上讨论的线性模型参数估计,都是假定未知参数之间是相互独立的,故对未知参数 X 没有附加任何约束条件,它是自由参数. 但在实际工作中,往往在未知参数 X 之间并非相互独立,而是存在线性关系(当然也可能是存在非线性关系,在此仅讨论线性关系). 此时,线

性模型(1-1-1)式就变为

$$\begin{cases} L = BX + \Delta \\ CX + C_0 = 0 \\ \mathrm{E}(\Delta) = 0 \\ \mathrm{Var}(\Delta) = \sigma^2 Q_{LL} \end{cases} \tag{1-1-37}$$

用 V 和 $\hat{X}_{\mathrm{LS}}$ 代替上式中的 Δ 和 X 可得

$$\begin{cases} V = B\hat{X}_{\mathrm{LS}} - L \\ C\hat{X}_{\mathrm{LS}} + C_0 = 0 \end{cases} \tag{1-1-38}$$

求(1-1-38)式的最小二乘解，就是在满足约束条件 $C\hat{X}_{\mathrm{LS}} + C_0 = 0$ 的情况下，求 $V'PV$ 的极小值. 为此，按求条件极值的方法组成函数

$$\varphi = V'PV = 2K'(C\hat{X}_{\mathrm{LS}} + C_0) \tag{1-1-39}$$

将上式对 $\hat{X}_{\mathrm{LS}}$ 求偏导数，并令其为零，得

$$\frac{\partial \varphi}{\partial \hat{X}_{\mathrm{LS}}} = 2V'P\frac{\partial V}{\partial \hat{X}_{\mathrm{LS}}} + 2K'C = 0$$

由此得

$$B'PV + C'K = 0 \tag{1-1-40}$$

将(1-1-40)式与(1-1-38)式联立求解，消去 V 后，得

$$\left.\begin{aligned} & B'PB\hat{X}_{\mathrm{LS}} + C'K - B'PL = 0 \\ & C\hat{X}_{\mathrm{LS}} + C_0 = 0 \end{aligned}\right\} \tag{1-1-41}$$

将(1-1-41)式的第一式左乘 $C(B'PB)^{-1}$ 后，减去第二式，得

$$C(B'PB)^{-1}C'K - (C(B'PB)^{-1}B'PL + C_0) = 0$$

解得

$$K = (C(B'PB)^{-1}C')^{-1}(C(B'PB)^{-1}B'PL + C_0) \tag{1-1-42}$$

将(1-1-42)式代入(1-1-41)式的第一式，得参数 X 的最小二乘估计为

$$\begin{aligned} \hat{X}_{\mathrm{LS}} = & (B'PB)^{-1}\{I - C'(C(B'PB)^{-1}C')^{-1}C(B'PB)^{-1}\}B'PL \\ & - (B'PB)^{-1}C'(C(B'PB)^{-1}C')^{-1}C_0 \end{aligned} \tag{1-1-43}$$

六、线性模型参数最小二乘估计的统一理论

在线性模型(1-1-1)式中，若 Q_{LL} 为已知的奇异方阵，即 $|Q_{LL}| = 0$，则称之为奇异线性模型，即

$$L = BX + \Delta, \mathrm{E}(\Delta) = 0, \mathrm{Var}(\Delta) = \sigma^2 Q_{LL}, |Q_{LL}| = 0 \tag{1-1-44}$$

由前述知，线性模型(1-1-1)式的最小二乘估计就是寻求使

$$V'PV = V'Q^{-1}V = \min \tag{1-1-45}$$

的未知参数的估值. 但在奇异线性模型(1-1-44)中，Q_{LL} 的凯利逆 Q_{LL}^{-1} 不存在. 这就使前述的一般最小二乘理论不能应用. 此时，人们自然地会想到用 Q_{LL} 的广义逆 Q_{LL}^{-} 来代替 Q_{LL} 的凯利逆 Q_{LL}^{-1}. 这样就变成求解 $V'Q_{LL}^{-}V = \min$ 的问题. 但由于 Q_{LL}^{-} 不惟一，$V'Q_{LL}^{-}V$ 就不惟一，即极小值不惟一. 为了解决这个问题，印度统计学家 C. R. Rao 提出了一种有效方法，从而建立了线性模型参数最小二乘估计的统一理论. 这一统一理论的基本思想是，寻求一个矩

阵 M 代替观测值的协因数阵 Q_{LL}，然后解优化问题：

$$V'M^{-}V=\min \tag{1-1-46}$$

在求解优化问题(1-1-46)式之前，人们自然会问，M 矩阵是否存在？若 M 矩阵存在，它的形式是怎样的？Rao 已证明[3](陈希孺，王松桂，1987)可以取 $M=Q_{LL}+BGB'$，其中 G 为对称矩阵，且满足 $\mathrm{rk}(G)=\mathrm{rk}(Q_{LL}\mid B)$. 这样的 M 适于 $Q_{LL}\geqslant 0$ 的情况. 当 $Q_{LL}>0$ 时，取 $G=0$. 当 $|Q_{LL}|=0$ 时，G 的一个简单的选择是 $G=kI,k>0$. 当 $M=Q_{LL}+BGB'$，G 为对称矩阵，且 $\mathrm{rk}(G)=\mathrm{rk}(Q_{LL}\mid B)$ 时，$V'M^{-}V$、$B'M^{-}B$ 和 $B'M^{-}L$ 都与 M^{-} 的选择无关[3](证明见陈希孺，王松桂，1987). 于是求解优化问题(1-1-46)式，得

$$\hat{X}_{\mathrm{LSR}}=(B'M^{-}B)^{-}B'M^{-}L \tag{1-1-47}$$

$\hat{X}_{\mathrm{LSR}}$表示由 C. R. Rao 给出的最小二乘估计.

七、线性模型参数的稳健估计

线性模型最小二乘估计是线性最优无偏估计，这一结论是在假设 $\mathrm{E}(\Delta)=0$, $\mathrm{Var}(\Delta)=\sigma^2 Q_{LL}$下得到的. 但在实际中，由于观测向量 L 中往往不可避免地存在粗差(Gross Error)，这时，线性模型中的假设就得不到满足. 当观测值不满足线性模型的假设条件，即观测值遭到粗差污染时，最小二乘估计具有明显的负面影响，即最小二乘估计不具有抵抗粗差干扰的特性，单个观测值的偏差也可能导致最小二乘估计面目全非[5](杨元喜，1993). 为此，人们就试图寻求一类参数估计方法，使其具有下述特点：

(1) 在假定的观测分布模型下，估值应是最优或接近最优的；

(2) 假设的分布模型与实际的分布模型有较小差异时，估值受粗差的影响较小；

(3) 当假设分布模型与实际分布模型有较大偏差时，估值不受破坏性的影响[6][7](罗永光，王海云，1987；周江文，黄幼才等，1997).

前两个特点是要求稳健估计在假定模型正确或假定模型略有偏差时，对估计结果影响较小. 这就只需要模型近似正确. 这两个特点的代价是在模型正确时，估计结果并非最优，只是接近最优. 第三个特点保证了在相当坏的情况下，估计结果也不会变得太坏. 在线性模型参数估计中，具有如上特点的估计称为线性模型参数的稳健估计，简称稳健估计(Robust Estimation).

稳健估计是一种能抵抗粗差干扰的估计. 其实质就是牺牲最小二乘估计的最优性，达到抵抗粗差污染的目的. 下面扼要介绍稳健估计. 为推导公式方便，将误差方程(1-1-2)式改写为

$$V=B\hat{X}_{\mathrm{R}}-L=\begin{pmatrix}b_1\\b_2\\\vdots\\b_n\end{pmatrix}\hat{X}_{\mathrm{R}}-\begin{pmatrix}L_1\\L_2\\\vdots\\L_n\end{pmatrix} \tag{1-1-48}$$

式中，b_i 为设计矩阵B 的第 i 行；$\hat{X}_{\mathrm{R}}$ 表示未知参数 X 的稳健估计.

设第 i 个观测值的权为 p_i，则按 M 估计原理，未知参数 X 的稳健估计就是求解下列优化问题：

$$\sum_{i=1}^{n}p_i\rho(v_i)=\sum_{i=1}^{n}p_i\rho(b_i\hat{X}_{\mathrm{R}}-L_i)=\min \tag{1-1-49}$$

上式对 $\hat{X}_R$ 求导数，并令其为零. 同时记 $\varphi(v_i)=\frac{\partial\rho}{\partial v_i}$，则有

$$\sum_{i=1}^{n} p_i\varphi(v_i)b_i = 0 \tag{1-1-50}$$

令

$$\frac{\varphi(v_i)}{v_i}=W_i, \bar{P}_{ii}=P_iW_i \tag{1-1-51}$$

式中：W_i 称为权因子，$\bar{P}_{ii}$称为等价权.

于是(1-1-50)式可写为

$$B'\bar{P}V=0 \tag{1-1-52}$$

将(1-1-48)式代入(1-1-52)式，得

$$B'\bar{P}B\hat{X}_R-B'\bar{P}L=0 \tag{1-1-53}$$

由此可得

$$\hat{X}_R=(B'\bar{P}B)^{-1}B'\bar{P}L \tag{1-1-54}$$

由于等价权 $\bar{P}$ 的引入，使得(1-1-54)式既能抵抗粗差的污染，又保留了最小二乘估计的形式. 使最小二乘估计公式简单，计算方便的优点得到充分体现. 所以周江文等[7]称(1-1-54)式为抗差最小二乘估计.

由(1-1-51)式可以看出，权因子 W_i 是残差 v_i 的非线性函数. 为了使等价权更切合实际，需要通过迭代计算，以改善权因子.

八、线性模型参数的信息扩散估计

由前述知，当观测值服从正态分布假设成立时，线性模型参数的最小二乘估计是最优线性无偏估计. 但当观测值遭到粗差污染时，最小二乘估计具有明显的负面影响. 单个粗差都可使最小二乘估计面目全非. 为了解决这个问题，数理统计学家们提出了具有抵抗粗差能力的稳健估计. 稳健估计是在假设观测值服从污染分布的前提下，通过牺牲最优性来达到抵抗粗差的目的. 在实际的参数估计过程中，人们一般不是选用最小二乘估计，就是选用稳健估计. 由于不知道观测值究竟服从什么分布，所以不论选用哪种估计，都要担很大的风险[8]（王新洲，1999）. 例如，当采用最小二乘估计时，要担观测值不服从正态分布，估计结果受粗差污染的风险；当采用稳健估计时，要担观测值服从正态分布，而估计结果并非最优的风险. 根据担此风险的原因知，如果对任何一组观测值，在参数估计之前，就能估计出它所服从的具体分布，然后再根据此分布进行参数估计，就可避免上述风险. 信息扩散估计(Information Diffusion Estimation)就是这样避免上述风险的一种参数估计方法. “信息扩散估计是一种智能估计，即当观测值中不含粗差(错误或异常值)时，参数估计的结果与最小二乘估计一样，是最优无偏估计；当观测值中含有粗差时，该估计不仅能很好地抵御粗差的影响，而且不需要迭代，比现有的稳健估计方法都好”[9]（黄崇福，2002）.

1. 信息扩散原理与扩散估计

设 $W=\{w_1,w_2,\cdots,w_n\}$ 是知识样本，L 是基础论域，记 w_i 的观测值为 l_i. 再设 $x=\varphi(l-l_i)$，则当 W 非完备时，存在函数 $\mu(x)$，使 l_i 点获得的量值为 1 的信息可按 $\mu(x)$ 的量值扩散到 l 上去. 且扩散所得到的原始信息分布 $Q(l)=\sum_{j=1}^{n}\mu(x)=\sum_{j=1}^{n}\mu(\varphi(l-l_i))$ 能

更好地反映 W 在总体的规律.这一原理称为信息扩散原理[10](黄崇福,王家鼎,1995).

根据这一原理对母体概率密度函数的估计称为扩散估计.扩散估计的确切定义如下:

定义 1-1-1 设 $\mu(x)$为定义在$(-\infty,\infty)$上的一个波雷尔可测函数,$\Delta_n>0$ 为常数,则称

$$\hat{f}(l)=\frac{1}{n\Delta_n}\sum_{j=1}^{n}\mu\left(\frac{l-l_j}{\Delta_n}\right) \tag{1-1-55}$$

为母体概率密度函数 $f(l)$的一个扩散估计[10](黄崇福,王家鼎,1995).式中:$\mu(x)$称为扩散函数;Δ_n 称为窗宽.

$$x=\frac{l-l_j}{\Delta_n} \tag{1-1-56}$$

2.扩散函数 $\mu(x)$的确定

由(1-1-55)式知,扩散估计的关键是扩散函数 $\mu(x)$的具体形式.对于不同的 $\mu(x)$,可得到不同的扩散估计结果.在《模糊信息优化处理技术及其应用》[10](黄崇福,王家鼎,1995)中根据分子扩散理论,导出的正态扩散函数为

$$\mu(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{1-1-57}$$

将(1-1-57)式代入(1-1-55)式,顾及(1-1-56)式,得母体概率密度函数 $f(l)$的正态扩散估计为

$$\hat{f}(l)=\frac{1}{n\Delta_n}\sum_{j=1}^{n}\left\{\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{\left(\frac{l-l_j}{\Delta_n}\right)^2}{2\sigma^2}\right]\right\}=\frac{1}{nh\sqrt{2\pi}}\sum_{j=1}^{n}\exp\left[-\frac{(l-l_j)^2}{2h^2}\right] \tag{1-1-58}$$

式中:

$$h=\sigma\Delta_n \tag{1-1-59}$$

(1-1-59)式所确定的 h 称为标准正态扩散的窗宽.

3.经验窗宽

由(1-1-58)式知,母体概率密度函数 $f(l)$的正态扩散估计 $\hat{f}(l)$不仅与观测值 l_i、子样容量(观测值的个数)n 有关,而且还与标准正态扩散的窗宽 h 有关.当观测完成后,观测值 l_i 和观测值的个数 n 都是已知量,此时只有窗宽 h 未知.因此,要根据观测值 l_i 来估计母体概率密度函数$f(l)$,首先要确定标准正态扩散的窗宽 h.在《模糊信息优化处理技术及其应用》[10](黄崇福,王家鼎,1995)中根据择近原则导出的窗宽 h 的经验公式为

$$h=\frac{\alpha(b-a)}{n-1} \tag{1-1-60}$$

式中:$a=\min(l_i)$,$b=\max(l_i)$,$i=1,2,\cdots,n$;α 是 n 的函数,α 与 n 的关系[8](王新洲,1999)列于表 1-1-1.

表 1-1-1

n	3	4	5	6	7
α	0.849321800	1.273982782	1.698643675	1.336252561	1.445461208
n	8	9	10	11	12
α	1.395189816	1.422962345	1.416278786	1.420835443	1.420269570

续表

n	13	14	15	16	17
α	1.420698795	1.420669671	1.420693321	1.420692226	1.420693101
n	18	19	20	21	22
α	1.420693101	1.420693101	1.420693101	1.420693101	1.420693101

由表 1-1-1 知，当 $n \geqslant 17$ 时，$\alpha \equiv 1.420693101$.

4. 最优窗宽

经验窗宽具有计算简单的优点. 但研究表明，经验窗宽不适用于所有分布[11]（游扬声，2001）. 而且采用不同的窗宽 h，母体概率密度函数 $f(l)$ 的估计精度是不同的. 因此，需要在某种最优准则下求得最优窗宽. 我们在母体概率密度函数估计的均方误差最小这一准则下导出了最优窗宽的迭代公式[12]（王新洲，游扬声，2001）：

$$\begin{cases} h^0 = b - \dfrac{a}{n-1} \\ h_i^{k+1} = \dfrac{(\hat{f}^k(l_i)(h^k)^4}{2n\sqrt{\pi}(\hat{f}^k(l_i+h^k) - 2\hat{f}^k(l_i) + \hat{f}^k(l_i-h^k)^2}, \quad h_i^{k+1} \leqslant \dfrac{b-a}{3} \\ h^{k+1} = \dfrac{1}{n}\sum\limits_i h_i^{k+1} \\ \hat{f}^{k+1}(l) = \dfrac{1}{\sqrt{2\pi}nh^k}\sum\limits_i \exp\dfrac{-(l-l_i)^2}{2(h^k)^2} \end{cases} \tag{1-1-61}$$

为了考察最优窗宽的效果，我们模拟了 100 个服从标准正态的观测值，分别采用经验窗宽（1-1-60）式和最优窗宽（1-1-61）式，对母体概率密度函数 $f(l)$ 进行了估计；同时还模拟了 40 个服从卡方（chi-squared）分布的观测值，分别采用经验窗宽（1-1-60）式和最优窗宽（1-1-61）式，对母体概率密度函数 $f(l)$ 进行估计. 估计结果见图1-1-1和图 1-1-2 所示.

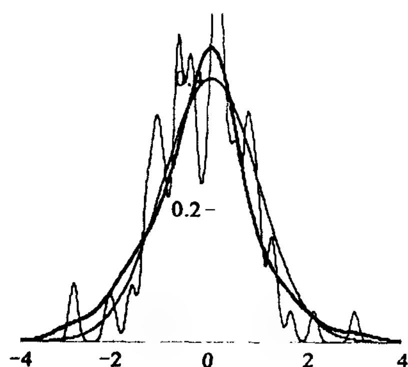


图 1-1-1　粗实线为最优窗宽估计结果；细实线为标准正态分布曲线；虚线为一般窗宽估计结果.

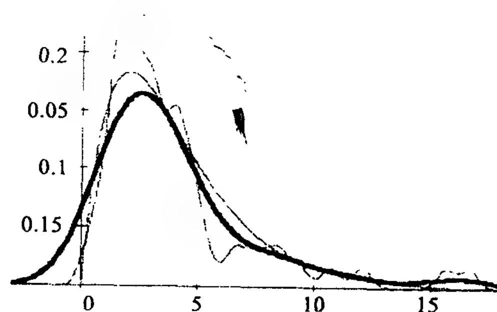


图 1-1-2　粗实线为最优窗宽估计结果；细实线为 chi-squared 分布曲线；虚线为一般窗宽估计结果.

由图 1-1-1 和图 1-1-2 可以看出，采用最优窗宽估计母体概率密度函数，其精度比采用

经验窗宽估计母体概率密度函数要高得多，并且其曲线能很好地逼近理论分布曲线.

5.单参数的信息扩散估计

单参数估计就是对于模型：

$$l_i=\mu+\Delta_i,\ i=1,2,\cdots,n \tag{1-1-62}$$

根据 n 个观测值 l_i 来估计母体的数学期望 μ.单参数的最小二乘估计就是算术平均值.其基本假设是观测值 l_i 来自正态母体.

单参数的信息扩散估计与最小二乘估计不同的是：不是先假设观测值的母体分布，而是在参数估计之前先根据观测向量 $l=(l_1,l_2,\cdots,l_n)$，应用(1-1-58)式估计出母体概率密度函数 $\hat{f}(l)$，然后再依据母体概率密度函数的估值 $\hat{f}(l)$ 来估计母体的数学期望 μ.王新洲[8](1999)导出的一维参数的信息扩散估计公式为

$$\hat{\mu}=\sum_{i=1}^{n}l_ip_i \tag{1-1-63}$$

式中：

$$p_i=\frac{\hat{f}(l_i)}{\sum_{i=1}^{n}\hat{f}(l_i)} \tag{1-1-64}$$

6.多参数的信息扩散估计

由(1-1-63)式定义的信息扩散估计，只适用于简单的线性模型(1-1-62)式.其原因是应用(1-1-58)式估计母体概率密度函数 $\hat{f}(l)$时，要求 n 个观测值 l_i 具有相同的数学期望.这一要求在测量数据处理中显然难以满足.因为测量数据处理中常用的线性模型是(1-1-1)式，而(1-1-1)式中的观测值 l_i 一般具有不同的数学期望 μ_i.

为了将单参数的信息扩散估计推广到一般的线性模型(1-1-1)式中去(即从单参数的信息扩散估计推广到多参数的信息扩散估计)，我们提出用标准化残差向量 W 作为扩散向量[13](游扬声，王新洲，2001)，即用标准化残差 W 的概率密度函数 $\hat{f}(w)$ 去代替观测值 l 的概率密度函数 $\hat{f}(l)$，近似地估计母体概率密度函数 $f(l)$.然后根据 $\hat{f}(w)$ 来确定观测值 l_i 的权 p_{l_i}.有了观测值 l_i 的权 p_{l_i} 后，对线性模型(1-1-1)式应用最小二乘估计即可.

多参数的信息扩散估计过程可描述为：

(1)令观测向量的权为单位矩阵，对线性模型(1-1-1)式进行最小二乘估计，并计算标准化残差向量 W.

(2)估计标准化残差向量 W 的概率密度函数 $\hat{f}(w)$.由(1-1-58)式得

$$\hat{f}(w)=\frac{1}{nh\sqrt{2\pi}}\sum_{j=1}^{n}\exp\left[-\frac{(w-w_j)^2}{2h^2}\right]$$

(3)用标准化残差向量 W 的概率密度函数 $\hat{f}(w)$ 代替观测值 l 的概率密度函数 $\hat{f}(l)$，并根据 $\hat{f}(w)$ 按(1-1-64)式确定观测值 l_i 的权 p_{l_i}：

$$p_{l_i}=\frac{\hat{f}(l_i)}{\sum_{i=1}^{n}\hat{f}(l_i)}=\frac{\hat{f}(w_i)}{\sum_{i=1}^{n}\hat{f}(w_i)} \tag{1-1-65}$$

(4)取观测向量的权为 $P=P(W)=\mathrm{diag}(p_{l_1},p_{l_2},\cdots,p_{l_n})$，再次对线性模型(1-1-1)式进行最小二乘估计，即得未知参数 X 的信息扩散估计：

$$\hat{X}=(B'P(W)B)^{-1}B'P(W)L \tag{1-1-66}$$

7. 信息扩散估计的抗差性

未知参数 X 的信息扩散估计(1-1-66)式，虽然实质上是最小二乘估计，但却兼有最小二乘估计和稳健估计的优点．所以，当观测值服从正态分布时，信息扩散估计和最小二乘估计一样，是未知参数 X 的最优线性无偏估计．当观测值含有粗差，不服从正态分布时，信息扩散估计具有很好的抵抗粗差的能力．这是因为在确定观测向量的权时考虑了观测值的分布信息，即按(1-1-65)式定权的结果．为什么按(1-1-65)式定权就具有抗差性呢？这是因为由(1-1-58)式和(1-1-65)式知，各观测值的权可写为

$$p_{l_i} = c_1 \sum_{j=1}^{n} \exp[-c_2(l_i - l_j)^2] \tag{1-1-67}$$

由(1-1-67)式知，当 l_i 含有粗差时，$(l_i - l_j)^2$ $(i \neq j)$ 必然较大，因此 $\sum_{j=1}^{n}\exp[-c_2(l_i - l_j)^2]$ 必然较小，从而导致 p_{l_i} 较小，这样就抑制了粗差的影响，使之具有抗差性．因此，当观测值含有粗差，不服从正态分布时，信息扩散估计优于最小二乘估计．为了验证这一点，我们应用信息扩散估计处理了武汉市轨道交通一号线一期工程的水准测量数据[14][15]（王新洲，2000 或 Wang Xinzhou，2001）．处理列于表 1-1-2．

表 1-1-2

水准点名	无粗差影响				有粗差影响			
	按距离定权		按分布定权		按距离定权		按分布定权	
	H_i(m)	σ_i(mm)	H_i(m)	σ_i(mm)	H_i(m)	σ_i(mm)	H_i(m)	σ_i(mm)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Wu15	23.6240		23.6240		23.6240		23.6240	
Wu16	22.6430		22.6430		22.6430		22.6430	
H1	24.6573	±1.4	24.6537	±2.4	24.6577	±3.7	24.6563	±2.1
H2	24.6251	±1.9	24.6223	±2.5	24.6259	±5.1	24.6245	±2.5
H5	25.0883	±2.4	25.0871	±2.5	25.0900	±6.2	25.0894	±2.7
H6	25.6422	±2.4	25.6408	±2.6	25.6433	±6.3	25.6423	±2.8
H4	24.1451	±2.4	24.1438	±2.8	24.1357	±6.3	24.1437	±3.7
H8	24.0928	±2.5	24.0919	±2.6	24.0909	±6.5	24.0926	±2.9
H9	26.1122	±2.4	26.1113	±2.6	26.1084	±6.4	26.1119	±3.0
GPS06	26.9144	±2.5	26.9136	±2.6	26.9120	±6.8	26.9142	±3.0
H13	25.4050	±2.4	25.4041	±2.5	25.4023	±6.4	25.4047	±2.9
GPS08	25.0220	±2.4	25.0214	±2.6	25.0190	±6.3	25.0217	±2.9
GPS11	23.4055	±2.5	23.4047	±2.5	23.4029	±6.5	23.4051	±2.9
H17	22.6157	±2.4	22.6150	±2.4	22.6132	±6.2	22.6153	±2.8
GPS12	24.2166	±2.3	24.2158	±2.4	24.2141	±6.2	24.2162	±2.8

续表

水准点名	无粗差影响				有粗差影响			
	按距离定权		按分布定权		按距离定权		按分布定权	
	H_i(m)	σ_i(mm)	H_i(m)	σ_i(mm)	H_i(m)	σ_i(mm)	H_i(m)	σ_i(mm)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GPS15	22.0897	±2.4	22.0895	±2.4	22.0874	±6.2	22.0894	±2.8
GPS17	21.2154	±2.2	21.2147	±2.2	21.2132	±5.8	21.2150	±2.5
GPS16	21.8294	±2.2	21.8287	±2.2	21.8270	±5.8	21.8290	±2.6
GPS19	24.7436	±2.2	24.7432	±2.1	24.7417	±5.8	24.7434	±2.5
GPS20	25.1099	±1.9	25.1098	±2.0	25.1082	±5.1	25.1098	±2.2
GPS27	24.3399	±1.9	24.3398	±1.9	24.3380	±5.1	24.3398	±2.2
GPS22	25.5464	±1.8	25.5468	+1.9	25.5454	±4.6	25.5467	±2.1
GPS24	25.0771	±1.3	25.0774	±1.6	25.0764	±3.4	25.0774	±1.7
GPS23	25.1108	±1.3	25.1110	±1.6	25.1088	±3.5	25.1110	±1.7
H31	23.6212	±0.8	23.6219	±1.3	23.6232	±2.1	23.6216	±1.4
GPS25	23.5229	±0.9	23.5234	±1.5	23.5182	±2.4	23.5232	±1.5
GPS21	24.2363	±1.8	24.2364	±1.9	24.2344	±4.8	24.2364	±2.1
H22	21.4305	±2.2	21.4300	±2.1	21.4285	±5.7	21.4302	±2.5
GPS14	21.2270	±2.4	21.2262	±2.4	21.2246	±6.2	21.2266	±2.8
GPS10	24.2806	±2.5	24.2804	±2.6	24.2779	±6.4	24.2804	±3.0
H11	26.5364	±2.5	26.5360	±2.6	26.5333	±6.4	26.5364	±3.0
H7	24.2187	±2.5	24.2176	±2.8	24.2113	±6.5	24.2178	±3.4

由表 1-1-2 知，当无粗差污染时，按分布定权的平差结果与按距离定权的平差结果（见表 1-1-2 的第 2、3 列和第 4、5 列）相差甚微．当我们在观测值中加上两个 10mm 的粗差时，按传统的平差方法（即按距离定权的最小二乘估计）的平差结果（见表 1-1-2 的第 6 列和第 7 列）受到粗差污染．由第 7 列可以看出，有粗差污染时各点的高程中误差是无粗差污染时的两倍以上．而按分布定权的平差结果（见表 1-1-2 的最后两列）仍然与无粗差污染时的平差结果相同．这表明信息扩散估计（即按分布定权的平差）具有较好的抗差性．同时也说明信息扩散估计具有合理性和实用性．

为了进一步说明信息扩散估计的抗差性能，我们将信息扩散估计与李德仁法、Huber 法等稳健估计方法进行了比较．结果发现当观测值中仅含一个粗差时，信息扩散估计与其他抗差估计方法一样，能很好地抵抗粗差．当观测值中含有多个粗差时，信息扩散估计明显优于其他方法．当观测值中含有$\frac{r}{2}$（r 为多余观测数）个粗差时，李德仁法和 Huber 法都会崩溃，而信息扩散估计的结果仍然很好．当粗差个数大于$\frac{r}{2}$时，信息扩散估计才发生崩溃．

九、线性模型参数的有偏估计

前面已经证明,线性模型(1-1-1)式的最小二乘估计 $\hat{X}_{LS}=(B'PB)^{-1}B'PL$ 具有最小方差.但是当设计矩阵 B 的列向量之间存在近似的线性关系时,法方程系数矩阵 $B'PB$ 接近奇异.称系数矩阵 $B'PB$ 接近奇异的法方程为病态方程.当法方程病态时,尽管 $\hat{X}_{LS}$的方差在线性无偏估计类中最小,但其数值却非常之大.即 $\hat{X}_{LS}$的精度比较差.其表现是法方程的解很不稳定.当 $B'PL$ 中有微小的误差时,通过$(B'PB)^{-1}$的作用,就会严重地放大该误差,使 $\hat{X}_{LS}$与 X 相差很大.为了在法方程病态时改进最小二乘估计,许多学者提出了一系列新的估计.其中很重要的一类估计就是有偏估计(Biased Estimation).有偏估计是牺牲最小二乘估计的无偏性,来换取比最小二乘估计更小的均方误差,从而达到改善 $\hat{X}_{LS}$的一类估计.在众多的有偏估计中,影响较大的是岭估计、广义岭估计、主成分估计和 Stein 压缩估计等[2](王松桂,1987).

1.岭估计

岭估计(Ridge Estimation)是 Hoerl 和 Kennard 1970 年提出来的.对于线性模型(1-1-1)式,其定义为

$$\hat{X}(k)=(B'PB+kI)^{-1}B'PL \tag{1-1-68}$$

式中:$k>0$ 为任意常数,称为岭参数.

岭估计具有如下性质:

(i)岭估计 $\hat{X}(k)$是最小的二乘估计的线性组合,由(1-1-68)式知

$$\begin{aligned}\hat{X}(k)&=(B'PB+kI)^{-1}B'PB(B'PB)^{-1}B'PL\\&=(B'PB+kI)^{-1}B'PB\hat{X}_{LS}\\&=\hat{X}_{LS}-k(B'PB+kI)^{-1}\hat{X}_{LS}\end{aligned} \tag{1-1-69}$$

(ii)岭估计 $\hat{X}(k)$是 X 的有偏估计,由(1-1-69)式得

$$\begin{aligned}\mathrm{E}[\hat{X}(k)]&=\mathrm{E}(\hat{X}_{LS})-k(B'PB+kI)^{-1}\mathrm{E}(\hat{X}_{LS})\\&=X-k(B'PB+kI)^{-1}X\end{aligned} \tag{1-1-70}$$

(iii)存在 $k>0$,使 $\hat{X}(k)$的均方误差小于 $\hat{X}_{LS}$的均方误差,即

$$\mathrm{MSE}[\hat{X}(k)]<\mathrm{MSE}(\hat{X}_{LS}) \tag{1-1-71}$$

由(1-1-68)式知,对于不同的 k 值,(1-1-68)式给出不同的估计.当 $k=0$ 时,$\hat{X}(0)=(B'PB)^{-1}B'PL=\hat{X}_{LS}$.即 $k=0$ 时的岭估计就是最小二乘估计.当设计矩阵 B 呈病态时,$B'PB$接近奇异,但 $B'PB+kI$ 接近奇异的程度得到改善.由于 $\hat{X}(k)$的均方误差小于 $\hat{X}_{LS}$的均方误差,所以,在 B' 呈病态时,岭估计改善了最小二乘估计.

2.广义岭估计

广义岭估计(Generalized Ridge Estimation)的定义为

$$X(k)=(B'PB+RKR')^{-1}B'PL \tag{1-1-72}$$

式中:R 为正交阵.$K=\mathrm{diag}(k_1,k_2,\cdots,k_t)$,是 t 个岭参数组成的对角阵.显然,当 $k_1=k_2=\cdots=k_t=k$ 时,(1-1-72)式变为(1-1-68)式.

3.Stein 估计

1955 年 Stein 证明了若法方程病态时,当 t 大于 2 时,最小二乘估计为不可容许估计,

即能找到另一个估计在均方误差意义下一致优于最小二乘估计. 人们称这种现象为 Stein 现象.

根据 Stein 现象, Stein 在 1955 年提出了通过压缩改进最小二乘的方法. 人们称 Stein 提出的方法为 Stein 估计. Stein 估计的定义为:

对于 $V=B\hat{X}_{\mathrm{LS}}-L$ 称

$$\hat{X}_{\mathrm{S}}(C)=C\hat{X}_{\mathrm{LS}} \tag{1-1-73}$$

为线性模型参数 X 的 Stein 估计, 其中 $0\leqslant C\leqslant 1$ 称为压缩系数.

Stein 估计为有偏估计, 即

$$\mathrm{E}(\hat{X}_{\mathrm{S}}(C))=C\mathrm{E}(\hat{X}_{\mathrm{LS}})=CX \tag{1-1-74}$$

当 $0<C<1$ 时, 有

$$\mathrm{MSE}(\hat{X}_{\mathrm{S}}(C))<\mathrm{MSE}(\hat{X}_{\mathrm{LS}}) \tag{1-1-75}$$

4. 在无偏估计类中改进最小二乘估计的方法

有偏估计实际上是在 $B'PB$ 病态时, 牺牲最小二乘估计的无偏性来对最小二乘估计进行改进. 在 $B'PB$ 病态时有偏估计虽然能改进最小二乘估计, 但估计结果有偏. 能否有办法既能改进最小二乘估计, 又能保证估计结果无偏呢? 笔者曾提出了这样的一种方法[16](王新洲, 1995). 下面简要介绍这种方法.

(1)约束条件

对于线性模型(1-1-1)式, 当 $B'PB$ 病态时, 最小二乘估计 $\hat{X}_{\mathrm{LS}}$很不可靠. 为了改善最小二乘估计, 对模型(1-1-1)式附加一个约束条件:

$$G'X=0 \tag{1-1-76}$$

式中: G' 为 $1\times t$ 的约束向量. 附加约束条件(1-1-76)后, 模型(1-1-1)式变为

$$\left.\begin{aligned}&L=BX+\Delta\\&G'X=0\\&\mathrm{E}(\Delta)=0\\&\mathrm{Var}(\Delta)=\sigma^2Q_{LL}\end{aligned}\right\} \tag{1-1-77}$$

其法方程为

$$\begin{aligned}&B'PB\hat{X}_G+GK-B'PL=0\\&G'\hat{X}_G=0\end{aligned} \tag{1-1-78}$$

法方程(1-1-78)的解为

$$\begin{aligned}&\hat{X}_G=Q_g(B'PL-GK)\\&K=(G'Q_gG)^{-1}G'Q_gB'PL\end{aligned} \tag{1-1-79}$$

式中:

$$Q_g=(B'PB+GG')^{-1}$$

令

$$C=1-Q_gG(G'Q_gG)^{-1}G' \tag{1-1-80}$$

则(1-1-79)式可写为

$$\hat{X}_G=CQ_gB'PL \tag{1-1-81}$$

(2)约束向量 G 的确定

在线性模型(1-1-1)式中附加约束条件(1-1-76)的目的, 是为了既改善法方程的态性, 又保证估计量无偏. 法方程的病态程度取决于 $B'PB$ 的条件数. $B'PB$ 的条件数由下式定义:

$$\mathrm{con}(B'PB)=\|B'PB\|\|(B'PB)^{-1}\| \tag{1-1-82}$$

当$0<\mathrm{con}(B'PB)<100$时，认为$B'PB$是良态的；当$100\leqslant\mathrm{con}(B'PB)\leqslant1000$时，认为$B'PB$有中等程度的病态；当$\mathrm{con}(B'PB)>1000$时，$B'PB$严重病态[3]（陈希孺，王松桂，1987）. 极端地，当$\mathrm{con}(B'PB)\to\infty$时，$B'PB$秩亏. 由此可知，对最小二乘估计的改进，应根据条件数来进行. 此外，$B'PB$的态性还与$B'PB$的行列式$|B'PB|$有关. 行列式$|B'PB|$的值越小，$B'PB$的病态越严重. 为此，我们根据行列式$|B'PB|$和条件数$\mathrm{con}(B'PB)$来确定约束向量G.

定义 1-1-2 约束向量G是下列线性方程的解：

$$(B'PB)G=|B'PB|e \tag{1-1-83}$$

式中：e为$t\times1$的向量，即$e=(1-\delta,1-\delta,\cdots,1-\delta)'$；

$$\delta=\begin{cases}0, & \mathrm{con}(B'PB)\geqslant100\\ 1, & \mathrm{con}(B'PB)<100\end{cases}$$

为了计算上的方便，按(1-1-83)式求得G后，再将其标准化，使下式成立：

$$G'G=I$$

分析(1-1-83)式知：

(i)当$\mathrm{con}(B'PB)<100$时，$\delta=1$，$e=(0,0,\cdots,0)'$. 于是(1-1-83)式变为齐次线性方程组. 由于$|B'PB|\neq0$，故此时(1-1-83)式没有非零解，只有零解，即$G=(0,0,\cdots,0)'$. 这表明，当$B'PB$为良态时，不需要附加约束条件. 即当$B'PB$为良态时，最小二乘估计不需要改进.

(ii)当$100\leqslant\mathrm{con}(B'PB)$时，$B'PB$不同程度地病态. 此时$\delta=0$，$e=(1,1,\cdots,1)'$. 由于$|B'PB|\neq0$，(1-1-83)式的常数项为$|B'PB|e$，故(1-1-83)式有惟一解：

$$G=(B'PB)^{-1}|B'PB|e=(B'PB)^{*}e \tag{1-1-84}$$

式中：$(B'PB)^{*}$为$B'PB$的伴随矩阵.

由于$e=(1,1,\cdots,1)'$，故(1-1-84)式可写为：

$$g_i=\sum_{j=1}^{t}(B'PB)_{ij}^{*} \tag{1-1-85}$$

(iii)当$\mathrm{con}(B'PB)\to\infty$时，有$|B'PB|=0$. 此时$\delta=0$，$e=(1,1,\cdots,1)'$. 由于$|B'PB|=0$，故(1-1-83)式的常数项为0. 此时(1-1-83)式又变为齐次线性方程组. 因为$|B'PB|=0$，故齐次线性方程组(1-1-83)式有非零解. 其非零解G就是$B'PB$的零特征值的特征向量. 这正好是秩亏自由网平差的情况.

由以上分析知，模型(1-1-77)式将法方程为良态、病态和秩亏时的平差问题统一起来了，也就是将最小二乘估计、岭估计和秩亏自由网平差统一起来了.

(3)$\hat{X}_G$的统计性质

$\hat{X}_G$具有无偏性，且$\hat{X}_G$的均方误差小于等于$\hat{X}_{\mathrm{LS}}$的均方误差，即

$$\mathrm{E}(\hat{X}_G)=CQ_gB'P\mathrm{E}(L)=CQ_gB'PBX=X$$

$$\mathrm{MSE}(\hat{X}_G)\leqslant\mathrm{MSE}(\hat{X}_{\mathrm{LS}})$$

详细证明见《在无偏估计类中改进最小二乘估计的方法》[16]（王新洲，1995）.

十、线性模型参数估计的谱修正迭代法

岭估计虽然能在某种程度上改善最小二乘估计，但它存在两个问题：第一，由于岭估计

改变了方程的等量关系，使得估计结果有偏；第二，岭参数 k 的确定非常困难，且随意性很大．那么，能否寻找一种算法，既改善法方程的态性，又不改变方程的等量关系，从而克服岭估计的两个缺点呢？在无偏估计类中改善最小二乘估计的方法，虽然能克服这两个缺点，但在线性模型(1-1-1)式中附加 $G'X=0$ 这一条件，没有充分的理由．为此，笔者又在对岭估计进一步研究的基础上，提出了一种新的迭代算法——谱修正迭代法[17]（王新洲，刘丁酉等，2001）．

1．修正迭代法简介

线性模型(1-1-1)式的法方程可写为

$$B'PB\hat{X}-B'PL=0 \tag{1-1-86}$$

将(1-1-86)式两边同时加上 $\hat{X}$，得

$$(B'PB+I)\hat{X}=B'PL+\hat{X} \tag{1-1-87}$$

式中：I 为 t 阶单位阵．

由于(1-1-87)式两边都有未知参数 $\hat{X}$，故只能采用迭代的方法求解(1-1-87)式．其迭代公式为

$$\hat{X}^{(k)}=(B'PB+I)^{-1}(B'PL+\hat{X}^{(k-1)}) \tag{1-1-88}$$

令

$$q=(B'PB+I)^{-1} \tag{1-1-89}$$

则(1-1-88)式可写为

$$\hat{X}^{(k)}=(q+q^2+\cdots+q^k)B'PL+q^k\hat{X}^{(0)} \tag{1-1-90}$$

式中 $\hat{X}^{(0)}$ 为未知参数 $\hat{X}$ 的初值．(1-1-88)式或者(1-1-90)式即为谱修正迭代法．

2．修正迭代法的收敛性

对谱修正迭代法(1-1-90)式，有：

定理 1-1-2 不论 $B'PB$ 呈良态、病态或秩亏，均有

$$\mathrm{rk}(B'PB+I)=t$$

即 $B'PB+I$ 为满秩矩阵．

定理 1-1-3 对于法方程(1-1-86)，当 $\mathrm{rk}(B'PB)=t$ 时，不论(1-1-86)式为良态或病态，谱修正迭代法(1-1-90)对任意的初值 $\hat{X}^{(0)}$，均有

$$\lim_{k\to\infty}\hat{X}^{(k)}=(B'PB)^{-1}B'PL=\hat{X}_{\mathrm{LS}} \tag{1-1-91}$$

即 $\hat{X}^{(k)}$ 收敛．

定理 1-1-4 对 $\mathrm{rk}(B'PB)=r<t$，即当法方程(1-1-86)式的系数矩阵秩亏时，谱修正迭代法(1-1-90)式对初值 $\hat{X}^{(0)}=0$，有

$$\lim_{k\to\infty}\hat{X}^{(k)}=(B'PB)^{+}B'PL \tag{1-1-92}$$

即 $\hat{X}^{(k)}$ 收敛．

为简化以上定理的证明，我们首先给出几条引理．

引理 1-1-1 设 P 为 n 阶正定阵，B 为 $n\times t$ 矩阵，则

1）当 $\mathrm{rk}(B)=t$ 时，$B'PB$ 为 t 阶正定阵，且其特征值

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_t>0; \tag{1-1-93}$$

2）当 $\mathrm{rk}(B)=r<t$ 时，$B'PB$ 为 t 阶半正定阵，且其特征值

$$\lambda_1\geqslant\cdots\geqslant\lambda_r>\lambda_{r+1}=\cdots=\lambda_t=0. \tag{1-1-94}$$

证明：由 P 的正定性，对 $\forall X \neq 0$，恒有 $X'NX = (BX)'P(BX) \geqslant 0$. 从而由 $B'PB$ 的正定性和半正定性知，存在 t 阶正交阵 $Q(QQ' = I)$，使 $B'PB = Q'DQ$，其中 $D = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_t\}$，且 Q' 的列向量为 $B'PB$ 的特征值所对应的单位正交特征向量. 引理 1-1-1 得证.

引理 1-1-2 设 B, P, I, q 如前所述，则

1) 当 $\mathrm{rk}(B'PB) = t$ 时，$\| q \|_2 < 1$；

2) 当 $\mathrm{rk}(B'PB) = r < t$ 时，$\| q \|_2 = 1$.

证明：因为 $q = (B'PB + I)^{-1}$，所以由引理 1-1-1 知，存在 t 阶正交阵 $Q(QQ' = I)$，使得

$$B'PB + I = Q'DQ + I = Q'(D + I)Q \tag{1-1-95}$$

这里 $D = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_t)$. 若记 $D_1 = \mathrm{diag}((\lambda_1 + 1)^{-1}, (\lambda_2 + 1)^{-1}, \cdots, (\lambda_t + 1)^{-1})$，则有

$$q = (B'PB + I)^{-1} = Q(D + I)^{-1}Q = Q'D_1Q$$

从而

1) 当 $\mathrm{rk}(B'PB) = t$ 时，$\| q \|_2 = \max(\lambda_{q'q}^{1/2}) = \max\limits_{1 \leqslant i \leqslant t}((\lambda_i + 1)^{-1}) = (\lambda_t + 1)^{-1} < 1$；

2) 当 $\mathrm{rk}(B'PB) = r < t$ 时，$\| q \|_2 = (\lambda_t + 1)^{-1} = 1$.

注：由上述证明可见，这里 $q = (B'PB + I)^{-1}$ 本质上是对 $(B'PB)^{-1}$ 或 $(B'PB)^{-}$ 的谱范数的一种修正，这就是我们为什么称算法(1-1-90)式为谱修正迭代法的理由.

引理 1-1-3 设 B, P, I, q 如前所述，则当 $\| q \|_2 < 1$ 时，矩阵幂级数

$$I + q + q^2 + \cdots + q^k + \cdots$$

绝对收敛，且其和为 $(1 - q)^{-1}$[18]（刘丁酉，1998）.

引理 1-1-4 设 B, P, I, q 如前所述，则当 $\| q \|_2 < 1$ 时，q 为幂收敛的[18]（刘丁酉，1998），即

$$\lim_{k \to +\infty} q^k = 0 \tag{1-1-96}$$

下面给出定理 1-1-2、1-1-3 和定理 1-1-4 的证明.

证明：由(1-1-95)式知，不论 $B'PB$ 呈良态、病态或秩亏，均有

$$|B'PB + I| = |Q'||D + I||Q| = (\lambda_1 + 1)(\lambda_2 + 1)\cdots(\lambda_T + 1) \geqslant 1$$

从而秩 $\mathrm{rk}(B'PB + I) = t$，即 $B'PB + I$ 为满秩矩阵. 定理 1-1-2 得证.

由引理 1-1-4 知，谱修正迭代法

$$\begin{aligned} \hat{X}^{(k)} &= (q + q^2 + \cdots + q^k)B'PL + q^k\hat{X}^{(0)} \\ &= (I + q + q^2 + \cdots + q^k)B'PL - B'PL \end{aligned}$$

再由引理 1-1-3 即知

$$\begin{aligned} \lim_{k \to \infty}\hat{X}^{(k)} &= (I + q + q^2 + \cdots + q^k + \cdots)B'PL - B'PL = [(I - q)^{-1} - I]B'PL \\ &= [(I - q)^{-1}(I - (I - q))]B'PL = [q^{-1}(I - q)]^{-1}B'PL \\ &= (q^{-1} - I)^{-1}B'PL \\ &= (B'PB)^{-1}B'PL = \hat{X}_{\mathrm{LS}} \end{aligned}$$

定理 1-1-3 得证.

为了证明定理 1-1-4，下面再给出几个引理.

引理 1-1-5 设 P 如前述，则存在非奇异阵 T，使 $P = T'T$.

证明： 因为 P 为 n 阶正定阵，所以存在正交阵 A，使得 $P = A'RA$，若记

$R = \mathrm{diag}(\mu_1, \mu_2, \cdots, \mu_n) = \mathrm{diag}(\mu_1^{1/2}, \mu_2^{1/2}, \cdots, \mu_n^{1/2}) \times \mathrm{diag}(\mu_1^{1/2}, \mu_2^{1/2}, \cdots, \mu_n^{1/2}) = R^{1/2}R^{1/2}$

于是

$$P = (R^{1/2}A)'(R^{1/2}A) = T'T$$

引理 1-1-6 设 P, B, Q 如前述，则[18]（刘丁酉，1998）：

$$B'P = Q'D^{1/2}T$$

证明： 因为 $B'PB = Q'D^{1/2}D^{1/2}Q = (D^{1/2}Q)'(D^{1/2}Q)$

又 $$N = B'PB = B'T'TB = (TB)'(TB)$$

比较上面两式，可知 $(TB)' = (D^{1/2}Q)'$，从而有：

$$B'P = (TB)'T = Q'D^{1/2}T$$

引理 1-1-7 设 B, P, N, q, U 如前述，则

$$q^kU = Q'\begin{bmatrix} (\lambda_1+1)^{-k} & & & \\ & \ddots & & \\ & & (\lambda_r+1)^{-k} & \\ & & & O_{(t-r)} \end{bmatrix}QU$$

证明： 因为

$$q = (N+I)^{-1} = [Q'(D+I)Q]^{-1} = Q'(D+I)^{-1}Q$$
$$U = B'PL = Q'D^{1/2}TL$$

所以

$$q^kU = Q'(D+E)^{-k}QQ'D^{1/2}TL$$

$$= Q'\begin{bmatrix} (\lambda_1+1)^{-k} & & & & & \\ & \ddots & & & & \\ & & (\lambda_r+1)^{-k} & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix}\begin{bmatrix} \lambda_1^{1/2} & & & & & \\ & \ddots & & & & \\ & & \lambda_r^{1/2} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}TL$$

$$= Q'\begin{bmatrix} (\lambda_1+1)^{-k} & & & & & \\ & \ddots & & & & \\ & & (\lambda_r+1)^{-k} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}\begin{bmatrix} \lambda_1^{1/2} & & & & & \\ & \ddots & & & & \\ & & \lambda_r^{1/2} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}TL$$

$$= Q'\mathrm{diag}((\lambda_1+1)^{-k}, (\lambda_2+1)^{-k}, \cdots, (\lambda_r+1)^{-k}, 0, \cdots, 0)QU$$

引理 1-1-8 设 $D = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r, 0, \cdots, 0)$，则[18]（刘丁酉，1998）：

$$D^+ = \mathrm{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \cdots, \lambda_r^{-1}, 0, \cdots, 0)$$

引理 1-1-9 设 P、B、N 如上所述，且 $\mathrm{rk}(N) = r < t$，则

$$N^+ = Q'D^+Q$$

证明：由 Moore-Penrose 逆的定义直接验证可得.

引理 1-1-10 记 $q_1 = Q'\mathrm{diag}\{(\lambda_1+1)^{-1},(\lambda_2+1)^{-1},\cdots,(\lambda_r+1)^{-1},0,\cdots,0\}Q$，则

$$\lim_{k\to\infty}(q_1+q_1^2+\cdots+q_1^k)=Q'\mathrm{diag}(\lambda_1^{-1},\lambda_2^{-1},\cdots,\lambda_r^{-1},0,\cdots,0)Q=(B'PB)^+ \tag{1-1-97}$$

证明：注意到$(\lambda_i+1)^{-1}+\cdots+(\lambda_i+1)^{-k}+\cdots=\dfrac{(\lambda_i+1)^{-1}}{1-(\lambda_i+1)^{-1}}=\lambda_i^{-1}$ 即证.

下面再来证明定理 1-1-4.

证明：由谱修正算法，并取 $\hat{X}^{(0)}=0$，则

$$\begin{aligned}\hat{X}^{(k)}&=(q+q^2+\cdots+q^k)B'PL\\&=qB'PL+q^2B'PL+\cdots+q^kB'PL\end{aligned}$$

再由引理 1-1-7 至引理 1-1-10 知

$$\begin{aligned}\hat{X}&=\lim_{k\to+\infty}\hat{X}^{(k)}=Q'\left(\sum_{k=1}^{\infty}\mathrm{diag}((\lambda_1+1)^{-1},\cdots,(\lambda_r+1)^{-1},0,\cdots,0)\right)QB'PL\\&=Q'\mathrm{diag}(\lambda_1^{-1},\cdots,\lambda_r^{-1},0,\cdots,0)QB'PL\\&=(B'PB)^+B'PL\end{aligned}$$

定理 1-1-4 得证.

3. 估计量 $\hat{X}$ 的协因数矩阵

全微分(1-1-90)式得

$$\mathrm{d}\hat{X}=\lim_{k\to\infty}(q+q^2+\cdots+q^k)B'P\mathrm{d}L \tag{1-1-98}$$

应用协因数传播律，得

$$Q_{\hat{X}\hat{X}}=\lim_{k\to\infty}(q+q^2+\cdots+q^k)(B'PB)(q+q^2+\cdots+q^k) \tag{1-1-99}$$

定理 1-1-5 当法方程系数矩阵 $B'PB$ 满秩时，有

$$Q_{\hat{X}\hat{X}}=(B'PB)^{-1} \tag{1-1-100}$$

当法方程系数矩阵 $B'PB$ 秩亏时，有

$$Q_{\hat{X}\hat{X}}=(B'PB)^{+} \tag{1-1-101}$$

下面来证明(1-1-100)式和(1-1-101)式.

证明：当法方程系数矩阵 $B'PB$ 满秩时，由引理 1-1-3 知

$$\lim_{k\to\infty}(I+q+q^2+\cdots+q^k)=(I-q)^{-1}$$

所以

$$\begin{aligned}\lim_{k\to\infty}(q+q^2+\cdots+q^k)&=(I-q)^{-1}-I\\&=(I-q)^{-1}(I-(I-q))\\&=(I-q)^{-1}q\\&=(q^{-1}(I-q))^{-1}\\&=(q^{-1}-I)^{-1}\\&=(B'PB)^{-1}\end{aligned} \tag{1-1-102}$$

将上式代入(1-1-99)式，得

$$Q_{\hat{X}\hat{X}}=(B'PB)^{-1}$$

(1-1-100)式得证.

为了证明(1-1-101)式,我们再给出下列引理.

引理 1-1-11 设 P、B 如上所述,且 $\mathrm{rk}(B'PB)=r<t$,则对于一切自然数 m、n,有

$$q^m B'PBq^n=(q_1)^m B'PB(q_1)^n \tag{1-1-103}$$

证明: 因为

$$q=(B'PB+I)^{-1}=(Q'(D+I)Q)^{-1}=Q'(D+I)^{-1}Q$$

所以

$$q^m=Q'(D+I)^{-m}Q,\ q^n=Q'(D+I)^{-n}Q.$$

于是

$$\begin{aligned}
q^m B'PBq^n &= Q'(D+I)^{-m}QQ'DQQ'(D+I)^{-n}Q\\
&= Q'\begin{bmatrix}(\lambda_1+1)^{-m} &&&&&\\ &\ddots&&&&\\ &&(\lambda_r+1)^{-m}&&&\\ &&&1&&\\ &&&&\ddots&\\ &&&&&1\end{bmatrix}\begin{bmatrix}\lambda_1 &&&&&\\ &\ddots&&&&\\ &&\lambda_r&&&\\ &&&0&&\\ &&&&\ddots&\\ &&&&&0\end{bmatrix}\cdot\\
&\quad\begin{bmatrix}(\lambda_1+1)^{-n} &&&&&\\ &\ddots&&&&\\ &&(\lambda_r+1)^{-n}&&&\\ &&&1&&\\ &&&&\ddots&\\ &&&&&1\end{bmatrix}Q\\
&= Q'\mathrm{diag}\{(\lambda_1+1)^{-m},\cdots,(\lambda_r+1)^{-m},0,\cdots,0\}QB'PBQ'\mathrm{diag}\{(\lambda_1+1)^{-n},\cdots,\\
&\quad(\lambda_r+1)^{-n},0,\cdots,0\}Q\\
&= q_1^m B'PBq_1^n
\end{aligned}$$

引理 1-1-12 设 B 为 t 阶常矩阵,$\{A_k\}_{k\geq 1}$ 为 t 阶矩阵序列,且 $\lim\limits_{k\to\infty}A_k=A$,则[18](刘丁酉,1998):

$$\lim_{k\to\infty}(A_kBA_k)=ABA \tag{1-1-104}$$

下面再来证明(1-1-101)式. 由引理 1-1-4 至引理 1-1-6 知,当 $B'PB$ 秩亏时,有:

$$\begin{aligned}
Q_{XX} &= \lim_{k\to\infty}\{(q+q^2+\cdots+q^k)B'PB(q+q^2+\cdots+q^k)\}\\
&= \lim_{k\to\infty}(qB'PBq+qB'PBq^2+\cdots+qB'PBq^k+q^2B'PBq+\cdots+q^kB'PBq^k)\\
&= \lim_{k\to\infty}(q_1B'PBq_1+q_1B'PBq_1^2+\cdots+q_1B'PBq_1^k+q_1^2B'PBq_1+\cdots+q_1^kB'PBq_1^k)\\
&= \lim_{k\to\infty}\{(q_1+q_1^2+\cdots+q_1^k)B'PB(q_1+q_1^2+\cdots+q_1^k)\}\\
&= (B'PB)^+(B'PB)(B'PB)^+\\
&= (B'PB)^+
\end{aligned}$$

(1-1-101)式得证.

4. 估计量 $\hat{X}$ 及其协因数矩阵 $Q_{\hat{X}\hat{X}}$ 的迭代程序

由于谱修正迭代法是通过迭代求解，而估计量 $\hat{X}$ 的协因数矩阵 $Q_{\hat{X}\hat{X}}$ 也需迭代求得，故可在求解估计量 $\hat{X}$ 的同时求得它的协因数矩阵 $Q_{\hat{X}\hat{X}}$. 其迭代程序如下：

(1)计算 $q=(B'PB+I)^{-1}$，并令 $M=q$, $Y=U$.

(2)计算 $M=q+Mq$, $X=MU$.

(3)如果 $|x_i-y_i|>\varepsilon$ ($i=1,2,\cdots,t$, ε 是充分小的正数，称为迭代误差限)，则 $Y=X$，转(2).

(4)计算 $S=MB'PBM$，输出 X 和 S.

S 就是估计量 $\hat{X}$ 的协因数矩阵 $Q_{\hat{X}\hat{X}}^{*}$

5.算例

例 1-1-1 本例取自《测量平差基础》[19](於宗俦，鲁林成，1983)上的例 6-12. 法方程为

$$\begin{pmatrix} 94.61 & -22.11 & -11.45 & -6.96 \\ -22.11 & 70.51 & -6.95 & -8.42 \\ -11.45 & -6.95 & 96.09 & -20.21 \\ -6.96 & -8.42 & -20.21 & 66.63 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -43.52 \\ 178.81 \backslash = -120.11 \\ -30.07 \end{pmatrix}$$

法方程系数矩阵 $B'PB$ 的特征值为：$\lambda=(112.5549 \quad 102.6065 \quad 71.2172 \quad 41.4614)'$

法方程系数矩阵 $B'PB$ 的条件数为：$\text{con}(N)=\dfrac{\lambda_{\max}}{\lambda_{\min}}=\dfrac{112.5549}{41.4614}=2.7147<100$.

所以该法方程为良态法方程. 该法方程按以上迭代程序迭代 3 次，得

$$\hat{X}^{(3)}=\begin{pmatrix} -0.1030 \\ 2.3209 \\ -1.2066 \\ -0.5348 \end{pmatrix},\quad Q_{\hat{X}^{(3)}\hat{X}^{(3)}}=\begin{pmatrix} 0.0120 & 0.0043 & 0.0023 & 0.0025 \\ 0.0043 & 0.0161 & 0.0024 & 0.0032 \\ 0.0023 & 0.0024 & 0.0117 & 0.0041 \\ 0.0025 & 0.0032 & 0.0041 & 0.0169 \end{pmatrix}$$

可见与高斯约化的结果完全相同.

例 1-1-2 本例取自王新洲、刘丁酉等(2001)的例 2. 法方程为

$$\begin{pmatrix} 2.1220 & -0.0269 & 0.0048 & 0.0019 \\ -0.0269 & 1.9949 & -0.0432 & -0.0168 \\ 0.0048 & -0.0432 & 0.8892 & 1.0178 \\ 0.0019 & -0.0168 & 1.0178 & 1.1656 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2.090033 \\ 3.765983 \\ 6.656877 \\ 7.684058 \end{pmatrix}$$

法方程系数矩阵 $B'PB$ 的特征值为

$$\lambda=(2.12992 \quad 2.07078 \quad 1.97089 \quad 0.00004)'$$

法方程系数矩阵 $B'PB$ 的条件数为

$$\text{con}(N)=\frac{\lambda_{\max}}{\lambda_{\min}}=\frac{2.12992}{0.00004}=53248>1000$$

所以该法方程为病态方程. 设未知参数的真值为 $X=(1 \quad 2 \quad 3 \quad 4)'$，由此真值算得的常数项为

---

* 当法方程为病态矩阵时，如此迭代有限次得到的 S 只是 $\hat{X}$ 的协因数矩阵 $Q_{\hat{X}\hat{X}}$ 的近似值. 其近似程度取决于迭代次数，迭代次数越多，近似程度越高.

$$B'PL=(2.090033 \quad 3.765983 \quad 6.656877 \quad 7.684058)'.$$

取 $B'PL=(2.090 \quad 3.765 \quad 6.656 \quad 7.684)'$，得

$\hat{X}=(B'PB)^{-1}B'PL=(1.0310 \quad 1.6570 \quad -20.9743 \quad 24.9290)'$，$\|\hat{X}-X\|_2=31.8262$

按以上迭代程序迭代 12 次，得

$$\hat{X}^{(12)}=\begin{pmatrix}0.9996\\2.0037\\3.2636\\3.7697\end{pmatrix},\quad Q_{\hat{X}^{(12)}\hat{X}^{(12)}}=\begin{pmatrix}0.4713 & 0.0063 & -0.0003 & -0.0010\\0.0063 & 0.5015 & 0.0026 & 0.0111\\-0.0003 & 0.0026 & 0.2131 & 0.2390\\-0.0010 & 0.0111 & 0.2390 & 0.2784\end{pmatrix}$$

$\|\hat{X}^{(10)}-X\|_2=0.3415$

因为 $\|\hat{X}^{(10)}-X\|_2=0.3415 \ll \|\hat{X}-X\|_2=31.8262$，可见谱修正迭代法能很好地改善最小二乘估计.

例 1-1-3 本例取自《测量平差基础》[19]（於宗俦，鲁林成，1983）上的例 7-5. 秩亏自由网平差的法方程为

$$\begin{pmatrix}5 & -2 & -1 & 2\\-2 & 5 & -1 & -2\\-1 & -1 & 3 & -1\\-2 & -2 & -1 & 5\end{pmatrix}\hat{X}=\begin{pmatrix}-11\\10\\-2\\3\end{pmatrix}$$

按以上迭代程序迭代 6 次，得

$$\hat{X}^{(6)}=\begin{pmatrix}-1.5000\\1.5000\\-0.5000\\0.5000\end{pmatrix},\quad Q_{\hat{X}^{(6)}\hat{X}^{(6)}}=\begin{pmatrix}0.1161 & -0.0268 & -0.0625 & -0.0268\\-0.0268 & 0.1161 & -0.0625 & -0.0268\\-0.0625 & -0.0625 & 0.1875 & -0.0625\\-0.0268 & -0.0268 & -0.0625 & 0.1161\end{pmatrix}$$

可见与《测量平差基础》[19]（於宗俦，鲁林成，1983）中的结果完全相同.

十一、线性模型参数的贝叶斯估计

上述各种估计，都是认为在观测之前，人们对参数 X 一无所知. 故 X 为非随机变量. 贝叶斯(Bayes)估计则不同，认为在观测之前，人们根据以往的经验或某种理论，就对参数 X 有了一些认识. 由于这些认识是在试验以前就有的，故称为先验信息. 表述这种先验信息的最简单和最方便的方法，就是将参数 X 看做随机量，并给出它的分布 $\pi(X)$. 这个分布一般称为参数 X 的先验分布[1]（成平，陈希孺等，1985）. 在对参数 X 进行估计时，充分顾及 X 的先验信息，这就是贝叶斯估计的最大特点.

下面扼要给出在平方损失下，观测向量 L 和参数 X 均服从正态分布时的贝叶斯估计公式.

设线性模型(1-1-1)式中的观测误差 Δ 服从正态分布 $N(0,D_\Delta)$，参数 X 的先验分布 $\pi(X)$ 也服从正态分布 $N(\mu,D_{X_0})$，则参数 X 的贝叶斯估计为

$$\hat{X}_B=\mu+D_{X_0}B'(BD_{X_0}B'+D_\Delta)^{-1}(L-B\mu) \tag{1-1-105}$$

因为

$$\mu + D_{X_0}B'(BD_{X_0}B' + D_\Delta)^{-1}(L - B\mu)$$
$$= \mu + D_{X_0}B'(BD_{X_0}B' + D_\Delta)^{-1}L - D_{X_0}B'(BD_{X_0}B' + D_\Delta)^{-1}B\mu$$
$$= \mu + D_{X_0}B'(BD_{X_0}B' + D_\Delta)^{-1}L + D_{X_0}B'D_\Delta^{-1}L$$
$$- D_{X_0}B'(BD_{X_0}B' + D_\Delta)^{-1}(BD_{X_0}B' + D_\Delta)D_\Delta^{-1}L - D_{X_0}B'(BD_{X_0}B' + D_\Delta)^{-1}B\mu$$
$$= \mu + D_{X_0}B'D_\Delta^{-1}L - D_{X_0}B'(BD_{X_0}B' + D_\Delta)^{-1}(BD_{X_0}B' + D_\Delta - D_\Delta)D_\Delta^{-1}L$$
$$- D_{X_0}B'(BD_{X_0}B' + D_\Delta)^{-1}B\mu$$
$$= \mu + D_{X_0}B'D^{-1}L - D_{X_0}B'(BD_{X_0}B' + D_\Delta)^{-1}BD_{X_0}B'D_\Delta^{-1}L$$
$$- D_{X_0}B'(BD_{X_0}B' + D_\Delta)^{-1}BD_{X_0}B'D_{X_0}^{-1}\mu$$
$$= D_{X_0}(B'D_{X_0}^{-1}L + D_{X_0}^{-1}\mu) - D_{X_0}B'(BD_{X_0}B' + D_\Delta)^{-1}BD_{X_0}(B'D_{X_0}^{-1}L + D_{X_0}^{-1}\mu)$$
$$= [D_{X_0} - D_{X_0}B'(BD_{X_0}B' + D_\Delta)^{-1}BD_{X_0}](B'D_{X_0}^{-1}L + D_{X_0}^{-1}\mu)$$
$$= (D_{X_0}^{-1} + B'D_{X_0}^{-1}B)^{-1}(B'D_{X_0}^{-1}L + D_{X_0}^{-1}\mu)$$

又因为 $D_{X_0}^{-1} = \frac{1}{\sigma^2}Q_{X_0}^{-1}, D_\Delta^{-1} = \frac{1}{\sigma^2}Q_{LL}^{-1}$，取 $\sigma^2 = 1$，则

$$D_{X_0}^{-1} = Q_{X_0}^{-1} = P_{X_0},\ D_{X_0}^{-1} = Q_{LL}^{-1} = P \tag{1-1-106}$$

所以参数 X 的贝叶斯估计又可写为[20]（张金槐，唐雪梅，1989）：

$$\hat{X}_B = (P_{X_0} + B'PB)^{-1}(B'PL + P_{X_0}\mu) \tag{1-1-107}$$

上面经过复杂的推导，才从(1-1-105)式推出我们所熟悉的(1-1-107)式. 事实上，按以下推导很容易得到(1-1-107)式.

因为贝叶斯估计 $\hat{X}_B$ 的先验信息已知，若令 $V_B = \hat{X}_B - \mu$，则有

$$V = B\hat{X}_B - L$$
$$V_B = \hat{X}_B - \mu \tag{1-1-108}$$

相应的权矩阵为

$$P = \begin{bmatrix} P & 0 \\ 0 & P_{X_0} \end{bmatrix} \tag{1-1-109}$$

在 $V'PV + V_B'PV_B = \min$ 下可得

$$\hat{X}_B = (P_{X_0} + B'PB)^{-1}(B'PL + P_{X_0}\mu)$$

当在观测前对参数 X 一无所知时，μ 可取任意实向量. 由于 μ 为任意实向量，则精度就非常之低，即 $D_{X_0} = \infty$. 于是有 $P_{X_0} = 0$. 将 $P_{X_0} = 0$ 代入(1-1-107)式，得无先验信息时的贝叶斯估计为

$$\hat{X}_B = (B'PB)^{-1}B'PL = \hat{X}_{LS} \tag{1-1-110}$$

(1-1-110)式表明，对于正态观测向量，当无参数 X 的先验信息时，线性模型参数的贝叶斯估计就是最小二乘估计.

十二、广义线性概括模型参数估计

在线性模型参数的最小二乘估计中，未知参数 X 为非随机参数. 而在线性模型参数的贝叶斯估计中，未知参数 X 为随机参数. 当线性模型中的未知参数一部分为随机参数，一部

分为非随机参数时，可用最小二乘配置来估计未知参数. 为了将各种线性模型统一在一个模型下面，笔者提出了广义线性概括模型[21]（王新洲，2000）. 现扼要介绍如下：

1. 函数模型

广义线性概括模型的函数模型定义为

$$\left.\begin{aligned} L &= BX + AY + \Delta \\ CX + C_0 &= 0 \end{aligned}\right\} \tag{1-1-111}$$

式中：L 为 $n \times 1$ 的观测值向量；A 为 $n \times m$ 的已知系数矩阵，其中 $A = (A_1 \quad 0)$，A_1 为 $n \times m_1$ 的已知系数矩阵，0 为 $n \times m_2$ 的矩阵；B 为 $n \times u$ 的已知系数矩阵；C 为 $d \times u$ 的已知系数矩阵；C_0 为 $d \times 1$ 的常数向量；X 为 $u \times 1$ 的非随机参数向量；Y 为 $m \times 1$ 的随机参数向量，且 $Y = \begin{pmatrix} S \\ S^w \end{pmatrix}$；$S$ 为 $m_1 \times 1$ 的已测点随机参数；S^w 为 $m_2 \times 1$ 的未测点随机参数；Δ 为 $n \times 1$ 的观测误差向量；n 为观测值的个数；u 为非随机参数的个数，$u \geqslant t$；m 为随机参数的个数，且 $m_1 + m_2 = m$；$d = u - t$ 为不独立的非随机参数的个数.

2. 随机模型

取单位权方差 $\sigma^2 = 1$，则

$$E(\Delta) = 0,\ E(Y) = \begin{pmatrix} E(S) \\ E(S^w) \end{pmatrix} = \begin{pmatrix} \mu_S \\ \mu_{S^w} \end{pmatrix},\ \mathrm{Var}(\Delta) = Q_{\Delta\Delta} = P_\Delta^{-1}$$

$$\mathrm{Var}(Y) = Q_{yy} = \begin{pmatrix} Q_{SS} & Q_{SS^w} \\ Q_{S^wS} & Q_{S^wS^w} \end{pmatrix} = P_y^{-1}$$

$$\mathrm{Var}(L) = Q_{LL} = P^{-1} = (Q_{\Delta\Delta} + B_1 Q_{SS} B_1'),\ \mathrm{Cov}(\Delta, Y) = 0,\ \mathrm{Cov}(Y, \Delta) = 0$$

3. 估计公式

令

$$L_y = \begin{pmatrix} L_S \\ L_{S^w} \end{pmatrix} = \mathrm{E}(Y) = \begin{pmatrix} \mu_S \\ \mu_{S^w} \end{pmatrix} \tag{1-1-112}$$

则虚拟观测方程为

$$L_y = Y + \Delta_y = \begin{pmatrix} S \\ S^w \end{pmatrix} + \begin{pmatrix} \Delta_S \\ \Delta_{S^w} \end{pmatrix} \tag{1-1-113}$$

于是由式(1-1-111)和式(1-1-113)可得误差方程和条件方程：

$$\left.\begin{aligned} V_y &= \hat{Y} - L_y \\ V &= B\hat{X} + A\hat{Y} - L \\ C\hat{X} + C_0 &= 0 \end{aligned}\right\} \tag{1-1-114}$$

再令

$$\bar{L} = \begin{pmatrix} L_y \\ L \end{pmatrix},\ \bar{V} = \begin{pmatrix} V_y \\ V \end{pmatrix},\ \hat{Z} = \begin{pmatrix} \hat{X} \\ \hat{Y} \end{pmatrix} = \begin{pmatrix} \hat{X} \\ \hat{S} \\ \hat{S}^w \end{pmatrix},\ \bar{\Delta} = \begin{pmatrix} \Delta_y \\ \Delta \end{pmatrix},\ \bar{B} = \begin{pmatrix} 0 & E \\ B & A \end{pmatrix},\ \bar{C} = (C \quad 0)$$

则有

$$\mathrm{Var}(\bar{\Delta}) = \begin{pmatrix} Q_{yy} & 0 \\ 0 & Q_{\Delta\Delta} \end{pmatrix},\ \bar{P} = \begin{pmatrix} Q_{yy}^{-1} & 0 \\ 0 & Q_{\Delta\Delta}^{-1} \end{pmatrix} = \begin{pmatrix} P_y & 0 \\ 0 & P_\Delta \end{pmatrix} \tag{1-1-115}$$

$$\left.\begin{aligned}\bar{V}&=\bar{B}\hat{Z}-\bar{L}\\ \bar{C}\hat{Z}+C_0&=0\end{aligned}\right\}\tag{1-1-116}$$

为了根据广义最小二乘原理

$$\bar{V}'\bar{P}\bar{V}=V'P_\Delta V+V_y'P_yV_y=\min\tag{1-1-117}$$

求解参数 $\hat{Z}$，组成函数

$$\Phi=\bar{V}'\bar{P}\bar{V}+2K'(\bar{C}\hat{Z}+C_0)\tag{1-1-118}$$

式中：K 为 $d\times 1$ 的联系数向量.

(1-1-118)式对 $\hat{Z}$ 求偏导数，并令其为零. 转置后得

$$\bar{B}'\bar{P}\bar{V}+\bar{C}'K=0\tag{1-1-119}$$

由(1-1-116)式和(1-1-119)式可得法方程

$$\left.\begin{aligned}\bar{B}'\bar{P}\bar{B}\hat{Z}+\bar{C}'K-\bar{B}'\bar{P}\bar{L}&=0\\ \bar{C}\hat{Z}+C_0&=0\end{aligned}\right\}\tag{1-1-120}$$

令

$$\bar{N}=\bar{B}'\bar{P}\bar{B},\ \bar{U}=\bar{B}'\bar{P}\bar{L}=\begin{pmatrix}B'P_\Delta L\\ P_yL_y+A'P_\Delta L\end{pmatrix}$$

则有

$$\begin{pmatrix}\bar{N}&\bar{C}'\\ \bar{C}&0\end{pmatrix}\begin{pmatrix}\hat{Z}\\ K\end{pmatrix}-\begin{pmatrix}\bar{U}\\ -C_0\end{pmatrix}=0\tag{1-1-121}$$

令

$$\bar{B}'\bar{P}\bar{B}=\begin{pmatrix}N_{11}&N_{12}\\ N_{21}&N_{22}\end{pmatrix}=\begin{pmatrix}B'P_\Delta B&B'P_\Delta A\\ A'P_\Delta B&P_y+A'P_\Delta A\end{pmatrix}$$

则(1-1-121)式可写为

$$\begin{bmatrix}N_{11}&N_{12}&C'\\ N_{21}&N_{22}&0\\ C&0&0\end{bmatrix}\begin{bmatrix}\hat{X}\\ \hat{Y}\\ K\end{bmatrix}\begin{bmatrix}B'P_\Delta L\\ P_yL_y+A'P_\Delta L\\ -C_0\end{bmatrix}\tag{1-1-122}$$

4. 各种特殊情况

(1)最小二乘配置

当 $u=t$ 时，即当(1-1-111)式中的非随机参数仅为 t 个独立的非随机参数时，有 $d=u-t=0$. 此时 $C=0$，$C_0=0$. 于是(1-1-111)式变为

$$L=BX+AY+\Delta\tag{1-1-123}$$

(1-1-123)式就是最小二乘配置的函数模型.

由于 $C=0$，$C_0=0$，所以(1-1-122)式变为

$$\begin{pmatrix}N_{11}&N_{12}\\ N_{21}&N_{22}\end{pmatrix}\begin{pmatrix}\hat{X}\\ \hat{Y}\end{pmatrix}=\begin{pmatrix}B'P_\Delta L\\ P_yL_y+A'P_\Delta L\end{pmatrix}\tag{1-1-124}$$

由 $u=t$ 知系数矩阵 B 列满秩. 于是式(1-1-124)的解为

$$\begin{pmatrix}\hat{X}\\ \hat{Y}\end{pmatrix}=\begin{pmatrix}N_{11}^{-1}+N_{11}^{-1}N_{12}R^{-1}N_{21}N_{11}^{-1}&-N_{11}^{-1}N_{12}R^{-1}\\ -R^{-1}N_{21}N_{11}^{-1}&R^{-1}\end{pmatrix}\begin{pmatrix}B'P_\Delta L\\ A'P_\Delta L+P_yL_y\end{pmatrix}\tag{1-1-125}$$

式中：
$$R = P_y + A'P_\Delta A - A'P_\Delta B(B'P_\Delta B)^{-1}B'P_\Delta A$$

即(见《测量平差基础》,武汉测绘科技大学测量平差教研室,1996)[4]：

$$\left.\begin{aligned}\hat{X} &= [B'(Q_{\Delta\Delta} + A_1Q_{SS}A_1')^{-1}B]^{-1}B'(Q_{\Delta\Delta} + A_1Q_{SS}A_1')^{-1}(L - A_1\mu_S)\\ \hat{Y} &= L_Y + Q_{YY}A'(Q_{\Delta\Delta} + AQ_{YY}A')^{-1}(L - B\hat{X} - AL_Y)\end{aligned}\right\} \tag{1-1-126}$$

顾及

$$A = (A_1 \quad 0),\ \hat{Y} = \begin{pmatrix}\hat{S}\\ \hat{S}^w\end{pmatrix},\ Q_{YY} = \begin{pmatrix}Q_{SS} & Q_{SS^w}\\ Q_{S^wS} & Q_{S^wS^w}\end{pmatrix},\ L_Y = \begin{pmatrix}\mu_S\\ \mu_{S^w}\end{pmatrix}$$

可得

$$\left.\begin{aligned}\hat{S} &= \mu_S + Q_{SS}A_1'(Q_{\Delta\Delta} + A_1Q_{SS}A_1')^{-1}(L - B\hat{X} - A_1\mu_S)\\ \hat{S}^w &= \mu_{S^w} + Q_{S^wS}A_1'(Q_{\Delta\Delta} + A_1Q_{SS}A_1')^{-1}(L - B\hat{X} - A_1\mu_S)\end{aligned}\right\} \tag{1-1-127}$$

(2)最小二乘滤波和推估

当 $u = 0$, 即当式(1-1-111)中不含非随机参数时, 有 $B = 0$, $C = 0$, $C_0 = 0$, 此时式(1-1-111)变为

$$L = AY + \Delta \tag{1-1-128}$$

由于 Y 为随机参数, 故(1-1-128)式就是最小二乘滤波和推估模型(武汉测绘科技大学测量平差教研室,1996)[4].

将 $B = 0$ 代入(1-1-126)式和(1-1-127)式, 得

$$\left.\begin{aligned}\hat{Y} &= L_Y + Q_{YY}A'(Q_{\Delta\Delta} + AQ_{YY}A')^{-1}(L - AL_Y)\\ \hat{S} &= \mu_S + Q_{SS}A_1'(Q_{\Delta\Delta} + A_1Q_{SS}A_1')^{-1}(L - A_1\mu_S)\\ \hat{S}^w &= \mu_{S^w} + Q_{S^wS}A_1'(Q_{\Delta\Delta} + A_1D_{SS}A_1')^{-1}(L - A_1\mu_S)\end{aligned}\right\} \tag{1-1-129}$$

由(1-1-129)式的第三式知, 虽然 S^w 为未测点上的随机参数, 且与观测值之间无函数关系, 但只要已知 S^w 和 S 的协因数矩阵 Q_{S^wS}, 就可以估计 S^w. 这说明事先了解各量之间的统计相关性是非常有用的.

(3)贝叶斯估计

在(1-1-122)式中, 当 $B = 0, C = 0, C_0 = 0$ 时, 有

$N_{11} = 0, N_{12} = 0, N_{21} = 0$, 顾及 $N_{22} = P_Y + A'P_\Delta A$, 则(1-1-122)式变为

$$(A'P_\Delta A + P_y)\hat{Y} = (A'P_\Delta L + P_yL_y) \tag{1-1-130}$$

于是有

$$\hat{Y} = (A'P_\Delta A + P_y)^{-1}(A'P_\Delta L + P_yL_y) \tag{1-1-131}$$

(1-1-131)式就是贝叶斯估计(1-1-110)式. 可见贝叶斯估计也是广义线性概括模型参数估计的特例.

(4)带线性约束的线性模型

当 $m = 0$, 即当式(1-1-111)中不含随机参数时, 有 $A = 0$, 此时式(1-1-111)变为

$$\left.\begin{aligned}L &= BX + \Delta\\ CX + C_0 &= 0\end{aligned}\right\} \tag{1-1-132}$$

(1-1-132)式就是前面介绍的带线性约束的线性模型. 相应的误差方程和条件方程为

$$\left.\begin{array}{l} V = B\hat{X} - L \\ C\hat{X} + C_0 = 0 \end{array}\right\}$$

由于 $m = 0$,所以 $P_y = 0, A = 0, P_\Delta = P_L = P$. 于是式(1-1-122)变为

$$\begin{pmatrix} B'PB & C' \\ C & 0 \end{pmatrix}\begin{pmatrix} \hat{X} \\ K \end{pmatrix} = \begin{pmatrix} B'PL \\ -C_0 \end{pmatrix}$$

其解为

$$\begin{pmatrix} \hat{X} \\ K \end{pmatrix} = \begin{pmatrix} B'PB & C' \\ C & 0 \end{pmatrix}^{-1}\begin{pmatrix} B'PL \\ -C_0 \end{pmatrix}$$

可见带线性约束的线性模型是广义线性概括模型的特例.

(5)一般的线性模型

当 $m = 0$,即当式(1-1-111)中不含随机参数 $A = 0$,且 $u = t$ 时,即当仅含 t 个独立的参数时,有 $d = u - t = t - t = 0$. 故 $C = 0, C_0 = 0$. 于是(1-1-111)式变为

$$L = BX + \Delta$$

这就是一般线性模型. 其解为

$$\hat{X}_{LS} = (B'PB)^{-1}B'PL$$

由以上推导知,模型(1-1-111)是包含最小二乘配置、最小二乘滤波和推估、贝叶斯估计、带线性约束的线性模型和一般线性模型的广义线性概括模型.

第二节 非线性模型参数估计问题的提出及其进展

从上节可以看到,线性模型参数估计理论十分成熟,成果异常之多. 但现实中的实际模型往往并不是线性模型. 而是非线性模型(Nonlinear Model). 用线性模型的理论来处理非线性模型,只是一种简单的、近似的方法. 这种近似往往带来很多问题,得出与事实不相符的结论. 因此,人们自然会想到,既然实际模型是非线性模型,就应该用非线性科学的方法来处理实际模型.

非线性科学在过去的30多年间激励了自然科学、工程技术与社会科学的几乎全部的科学研究人员,并向人们提出了划时代的挑战. 正是由于非线性的作用,才孕育出大自然的万千气象,人类社会的风云变幻和人类思维的错综差异[22](焦李成,1990). 近几十年来,耗散结构、孤粒子、自组织、协同学、超循环与微循环、奇异吸引子与混沌动力学等理论的问世和发展,使人们认识到非线性是一切复杂问题之源[22](焦李成,1990). 非线性科学就是研究复杂性现象的一类新学科. 因此,要深入地、精确地、符合实际地研究各种自然现象乃至社会现象,就必须研究非线性问题. 非线性问题涉及到自然科学的方方面面. 周光召1995年12月在《人民日报》发表的《迈向科技大发展的新世纪》一文中指出:"非线性科学是关于体系总体本质的一门新学科,它更着重于总体、过程和演化. 因此,透过这扇窗户,看到的将与牛顿、爱因斯坦等人所创建的决定性的、简单和谐的模式不同,而是一个演化的、开放的、复杂的世界,这是一幅更接近真实的世界图景."他同时还指出:"非线性科学不仅在认识论上有重大的哲学意义,在求解基本问题时有重大科学意义,而且在研究生态环境、医疗诊断、经济发展、科学决策等许多问题时,都有重要的应用价值."

由于非线性问题的理论研究远比线性问题的研究困难得多、复杂得多，所以许多国家专门成立了非线性问题研究机构，并设立研究课题. 例如，国家自然科学基金委员会采纳郝柏林等提出的建议，在"八五"期间，把非线性科学作为全国10项重大课题之一，列入攀登计划首批项目，由著名数学家谷超豪担任这一课题的首席科学家.

非线性科学在测绘界也同样受到重视. 国家自然科学基金委员会1994年在自然科学学科发展战略调整调查报告《大地测量学》(国家自然科学基金委员会，1994)[23]一书中，明确提出将非线性模型参数估计理论作为大地测量学科发展的重大基础理论问题之一. 这是因为就参数估计而言，线性模型中的很多理论在非线性模型中就不一定适用；线性模型中的很多结论在非线性模型中就不一定成立；线性模型参数估计中的很多优良统计性质在非线性模型中就不一定存在. 例如，在线性模型中，当随机误差服从正态分布时，未知参数 X 的最小二乘估计$\hat{X}_{LS}$具有一致无偏性和方差最小性. 但在非线性模型参数估计中，即使随机误差严格服从正态分布，未知参数 X 的非线性最小二乘估计$\hat{X}_{NLS}$也是有偏的. 其方差一般都不能达到CR下界. "非线性"本身是一种数字特征，但非线性问题即使在数学中也并不成熟[24](韦博成，1989). 因此，必须对非线性模型参数估计问题进行深入的研究.

现实世界中，严格的线性模型并不多见，它们或多或少都带有某种程度的近似. 随着科学技术和近代统计学的飞速发展，不能简单化为线性模型的非线性模型越来越多地出现在统计学家面前. 农业、生物、经济、工程技术等各部门都提出了许多非线性模型以及其他非线性统计问题. 因此，积极开展非线性模型参数估计的研究在理论与实践中日趋重要. 非线性模型参数估计是线性模型参数估计的自然推广，也是必然发展趋势[24](韦博成，1989). 另外，对线性模型参数估计的理论研究已臻完善，而对非线性模型参数估计的研究，理论还不成熟. 所以从理论上讲，有必要对非线性模型参数估计理论进行深入的研究.

在测量上，大量的数学模型也是非线性模型. 现代大地测量的观测方程多具有很强的非线性性，对非线性展开初值点十分敏感，需要研究模型空间的容许曲率问题，以及非线性函数空间的平差理论和可普及的适用方法(国家自然科学基金委员会，1994)[23]. 对于大地测量中大量的非线性模型，传统的做法是线性近似(测量上称之为线性化)，即将其展为泰勒级数，并取至一次项，略去二次以上各项. 如此线性近似，必然会引起模型误差. 过去由于测量精度不高，线性近似所引起的模型误差往往小于观测误差，故可忽略不计. 随着科学技术的不断发展，现在的观测精度已大大提高，致使因线性近似所产生的模型误差与观测误差相当，有些甚至还会大于观测误差. 例如，GPS载波相位观测值的精度很高，往往小于因线性近似所产生的模型误差.

鉴于上述各种原因，有必要结合测量数据处理的实际，对非线性模型参数估计进行深入的研究.

关于非线性模型参数估计理论的研究，始于20世纪60年代初期. 但其进展并不快. 直到1980年，加拿大统计学家Bates和Watts引入曲率度量以后，才得到较快的发展[24](韦博成，1989). 测量上对非线性模型参数估计理论的研究相对较晚. 较深入的研究是20世纪80年代后期的事. 1985年之后，国际著名的大地测量学者P.J.G.Teunissen在非线性模型参数估计方面作了卓有成效的研究[25]. 先后研究了非线性模型最小二乘估计的一、二阶矩. 阐述了非线性模型的识别、度量非线性强度的指标以及非线性模型曲率的几何意义等. 通过对展开式中舍去项造成函数模型偏差的研究，提出从舍去的项中找出其对函数模型和参数的影

响，然后对函数模型和参数的估值进行修正（P. J. G. Teunissen, 1985, 1988, 1989, 1990）[26]~[28]. Blaha系统地研究了非线性最小二乘的无迭代求解理论[29]（Blaha. G. 1994）. Lohse 系统地研究了非线性模型的参数估计理论[30]（Lohse. P. 1993）. Athanasios Dermanis 和 Fernando Sanso研究了可容许和不可容许的非线性估计原理，提出了非线性估计的贝叶斯方法[31]（Athanasios Dermanis and Fernando Sanso）.

在我国，广大测量学者曾对非线性模型参数估计中的一些问题作过零星的研究. 例如徐培亮研究了非线性函数的协方差传播公式[32]（徐培亮，1986）. 刘大杰、黄加纳研究过非线性最小二乘的迭代解法[33]（刘大杰，黄加纳，1987）. 周世健研究了广义方差-协方差传播[34]（周世健，1996）. 刘国林、陶华学在非线性模型参数估计方面做了一些系统的研究工作. 大量工作都是对非线性模型展开后取至二次项，然后再对其加以研究，他们研究了这种新非线性模型下的平差问题（如秩亏自由网平差）和协因数的传播问题[35][36]（刘国林，陶华学，1997, 2000）. 胡圣武、陶本藻对非线性模型参数估计的统计性质作了初步研究，并将其应用到 GIS 中[37][38]（胡圣武，陶本藻，1997）. 笔者在国家自然科学基金的资助下，对非线性模型参数估计理论作了系统的研究. 从非线性模型非线性强度的度量，到如何判断一个非线性模型能否线性近似；从非线性模型参数估计的算法，到非线性模型中单位权方差的估计都作了深入的研究. 提出了非线性模型线性近似时容许曲率的概念；给出了非线性模型能否线性近似的实用判据；导出了非线性模型参数估计的直接解法和非线性模型参数估计中单位权方差的估计公式[39]~[43]（王新洲，1997, 1999, 2000）. 李朝奎、徐望国等根据非线性模型参数估计应该不对非线性模型线性近似，而应该直接解算非线性目标函数这一宗旨，讨论了目标函数的常用算法——迭代法和非线性目标函数的最优算法——不依赖函数导数的多维轮环搜索法和基于差分理论的迭代算法[44]（李朝奎，徐望国等，2001）.

本书在全面总结这些研究成果的基础上，又增加了笔者最新的研究成果. 希望本书在测量数据处理方面能起到抛砖引玉的作用.

第二章　非线性模型非线性强度的度量

第一节　非线性强度的概念

在测量数据处理中,我们熟悉的线性观测方程的一般形式为

$L_i = b_{i1}x_1 + b_{i2}x_2 + \cdots + b_{it}x_t + b_{i0} + \Delta_i \quad (i = 1, 2, \cdots, n)$

其矩阵形式为

$$L = BX + B_0 + \Delta \tag{2-1-1}$$

式中:$L = (L_1 \quad L_2 \quad \cdots \quad L_n)'$为$n \times 1$的观测向量;$n$为观测值的个数;$X$为$t \times 1$的未知参数向量;$t$为必要观测的个数;$B_0 = (B_{01} \quad B_{02} \quad \cdots \quad B_{0n})'$为$n \times 1$的常数向量;$\Delta = (\Delta_1 \quad \Delta_2 \quad \cdots \quad \Delta_n)'$为$n \times 1$的观测误差向量;$B$为$n \times t$的设计矩阵,即

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1t} \\ b_{21} & b_{22} & \cdots & b_{2t} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nt} \end{pmatrix}$$

比如水准测量中,当不设尺度比参数,且以待定点高程为未知参数时,其观测方程就是(2-1-1)式所示的线性形式.然而,测量中更多的观测方程是非线性方程.比如导线测量中,以待定点坐标为未知参数的角度观测方程和边长观测方程分别为

$$\left.\begin{aligned} \beta_i &= \arctan\frac{y_k - y_j}{x_k - x_j} - \arctan\frac{y_h - y_j}{x_h - x_j} + \Delta_{\beta_i} \\ S_i &= \sqrt{(x_h - x_j)^2 + (y_h - y_j)^2} + \Delta_{S_i} \end{aligned}\right\} \tag{2-1-2}$$

它们都是待定点坐标(x_i, y_i)的非线性函数.又如在GPS伪距测量中,第j颗卫星至测站k的几何距离的观测方程为

$$\rho_k^j = \sqrt{(x_k - x^j)^2 + (y_k - y^j)^2 + (z_k - z^j)^2} + c\delta t + \Delta_k^j \tag{2-1-3}$$

也是测站点k的待定坐标(x_k, y_k, z_k)的非线性函数.一般地,用L表示$n \times 1$的观测向量,用X表示$t \times 1$的未知参数向量,用Δ表示$n \times 1$的观测误差向量,则非线性观测方程可写为

$$L = f(X) + \Delta \tag{2-1-4}$$

式中,$f(X) = (f_1(X) \quad f_2(X) \quad \cdots \quad f_n(X))'$,是由$n$个$X$的非线性函数组成的$n \times 1$的向量,(2-1-4)式就是我们所要讨论的一般非线性模型.

为了推导公式方便,不失一般性,设L为同精度独立观测值.

关于模型(2-1-4)的参数估计,传统的方法是将其线性化,即将(2-1-4)式在参数的近

似值 X_0 处展为泰勒级数，并仅取至一次项. 然后再应用线性模型参数估计理论进行参数估计. 将非线性模型(2-1-4)式线性化，因略去了二阶以及二阶以上各高次项，得到的仅含一次项的线性模型显然是原模型的近似模型. 因此，我们以后将传统线性化称为线性近似. 由于线性近似后得到的线性模型是近似模型，所以线性近似必然会产生模型误差. 从下面两个例子可清楚地看到这一点.

例 2-1-1 已知非线性模型为 $L_i = x_1 e^{ix_2}$. 其中参数 x_1 和 x_2 的真值为 $X = (5.420136187 \quad -0.25436189)'$. L_i 的 5 个真值(用参数的真值 X 算得)和相应的 5 个同精度独立观测值列于表 2-1-1.

表 2-1-1 **L_i 的真值和相应的观测值**

i	1	2	3	4	5
真值	4.202834	3.258924	2.527006	1.959469	1.519394
观测值	4.20	3.25	2.52	1.95	1.51

观测值的中误差 $\sigma_0 = \pm 0.007833$，观测方程为

$$L_1 = x_1 e^{x_2} + \Delta_1$$
$$L_2 = x_2 e^{2x_2} + \Delta_2$$
$$L_3 = x_1 e^{3x_2} + \Delta_3$$
$$L_4 = x_1 e^{4x_2} + \Delta_4$$
$$L_5 = x_1 e^{5x_2} + \Delta_5$$

取参数 X 的近似值为 $X_0 = (x_{01} \quad x_{02})' = (5.4 \quad -0.3)'$. 将观测方程在 X_0 处线性近似，得误差方程

$$V = B\mathrm{d}X + l = \begin{pmatrix} 0.7408 & 4.0004 \\ 0.5488 & 5.9272 \\ 0.4066 & 6.5864 \\ 0.3012 & 6.5058 \\ 0.2231 & 6.0245 \end{pmatrix} \begin{pmatrix} \mathrm{d}x_1 \\ \mathrm{d}x_2 \end{pmatrix} - \begin{pmatrix} 0.1996 \\ 0.2864 \\ 0.3245 \\ 0.3236 \\ 0.3051 \end{pmatrix}$$

根据最小二乘原理，得 $\mathrm{d}X$ 的最小二乘估计为

$$\mathrm{d}X = \begin{pmatrix} -0.005858021 \\ 0.049953787 \end{pmatrix}$$

于是，参数 X 的最小二乘估值为

$$\hat{X} = X_0 + \mathrm{d}X = \begin{pmatrix} 5.394141979 \\ -0.250246213 \end{pmatrix}$$

参数估值 $\hat{X}$ 的真误差为

$$\Delta X = \hat{X} - X = \begin{pmatrix} -0.025994200 \\ 0.004315680 \end{pmatrix}$$

$$\| \Delta X \| = 0.02631801$$

参数估值 $\hat{X}$ 的真误差ΔX 主要由两种误差引起. 一种为观测误差;另一种为线性近似所引起的模型误差.

由于观测值的中误差为 $\sigma_0 = \pm 0.007833$,而参数估值的协因数阵为

$$Q_{\hat{X}\hat{X}} = (B'B)^{-1} = \begin{pmatrix} 3.3738685 & -0.2377004 \\ -0.2377004 & 0.0225226 \end{pmatrix}$$

故由观测误差引起的参数估值的中误差分别为

$$\sigma_{\hat{x}_1} = \pm 0.014388, \quad \sigma_{\hat{x}_2} = \pm 0.001176$$

于是有$\dfrac{|\Delta x_1|}{|\sigma_{\hat{x}_1}|} = 1.807$, $\dfrac{|\Delta x_2|}{|\sigma_{\hat{x}_2}|} = 3.670$, 即参数估值的实际误差大约是其中误差的 2 倍. 这说明此例中由线性近似所产生的模型误差大于、等于由观测误差所引起的误差.

例 2-1-2 已知非线性模型为 $L_i = \dfrac{x_1^2}{i} + x_2$. 其中参数的真值仍为 $X = (5.420136187 \quad -0.25436189)'$, L_i 的 5 个真值和相应的 5 个同精度独立观测值列于表 2-1-2.

表 2-1-2 **L_i 的真值和相应的观测值**

i	1	2	3	4	5
真值	29.123514	14.434576	9.538264	7.090107	5.621213
观测值	29.12	14.43	9.53	7.09	5.62

观测值的中误差为 $\sigma_0 = \pm 0.004540$, 仍取参数的近似值为 $X_0 = (x_{01} \quad x_{02})' = (5.4 \quad -0.3)'$, 相应的观测方程为

$$L_1 = x_1^2 + x_2 + \Delta_1$$

$$L_2 = \frac{x_1^2}{2} + x_2 + \Delta_2$$

$$L_3 = \frac{x_1^2}{3} + x_2 + \Delta_3$$

$$L_4 = \frac{x_1^2}{4} + x_2 + \Delta_4$$

$$L_5 = \frac{x_1^2}{5} + x_2 + \Delta_5$$

在 X_0 处线性近似, 得 5 个误差方程为

$$V = B\mathrm{d}X + l = \begin{pmatrix} 10.8 & 1 \\ 5.4 & 1 \\ 3.6 & 1 \\ 2.7 & 1 \\ 2.16 & 1 \end{pmatrix} \mathrm{d}X - \begin{pmatrix} 0.26 \\ 0.15 \\ 0.11 \\ 0.10 \\ 0.088 \end{pmatrix}$$

$\mathrm{d}X$ 的最小二乘估计为

$$\mathrm{d}X=\begin{pmatrix}0.020007626\\0.042922386\end{pmatrix}$$

于是

$$\hat{X}=X_0+\mathrm{d}X=\begin{pmatrix}5.420007626\\-0.257077610\end{pmatrix}$$

$\hat{X}$ 的真误差为

$$\Delta X=\begin{pmatrix}\Delta x_1\\\Delta x_2\end{pmatrix}=\begin{pmatrix}-0.000128561\\-0.002715720\end{pmatrix}$$

$$\|\Delta X\|=0.00271876$$

显然,本例中 $\hat{X}$ 与其真值的距离比例 2-1-1 中 $\hat{X}$ 与其真值的距离要小一个数量级.

因为 $\hat{X}$ 的协因数阵为 $Q_{\hat{X}\hat{X}}=(B'B)^{-1}=\begin{pmatrix}0.020369718 & -0.10046345\\-0.10046345 & 0.695485744\end{pmatrix}$

所以参数估值的中误差为

$$\sigma_{\hat{x}_1}=\pm0.000648,\ \sigma_{\hat{x}_2}=\pm0.003786$$

$\frac{|\Delta x_1|}{|\sigma_{\hat{x}_1}|}=0.198$, $\frac{|\Delta x_2|}{|\sigma_{\hat{x}_2}|}=0.717$. 即参数估值的实际误差比中误差要小. 这说明此例中由线性近似所产生的模型误差可忽略不计.

例 2-1-1 和例 2-1-2 的观测精度基本相同,参数的真值和其相应的近似值也相同,但例 2-1-1 中参数估值的精度却远远低于例 2-1-2 中参数估值的精度. 其原因主要是例 2-1-1 线性近似时引起了较大的模型误差. 而例 2-1-2 线性近似时引起的模型误差较小,可忽略不计.

为什么不同的非线性模型线性近似时会引起不同的模型误差呢? 这是因为不同的非线性模型的"非线性"程度不一样. "非线性"程度越强,线性近似时产生的模型误差就越大. 非线性模型的"非线性"程度,称为非线性强度(Non-linearity). 显然,非线性强度越强,线性近似时产生的模型误差就越大. 因此,一个非线性模型,采用线性近似的方法进行参数估计时,参数估值的精度很大程度上取决于该模型的非线性强度.

第二节　非线性强度的度量

一、固有曲率和参数效应曲率

不同的非线性模型具有不同的非线性强度. 而非线性强度直接影响线性近似的效果. 为了评价线性近似的优劣程度,就需要定义一个数量指标来度量非线性强度. Beale 1960 年定义了四种曲率度量[24](韦博成, 1989),其中基本的一种为

$$\hat{N}=\frac{\|f(X)-f(\hat{X})-B(\hat{X})(X-\hat{X})\|^2}{\|f(X)-f(\hat{X})\|^2}\tag{2-2-1}$$

式中：

$$B(\hat{X})=\begin{bmatrix}\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_t}\\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_t}\\ \vdots & \vdots & & \vdots\\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_t}\end{bmatrix}_{X=\hat{X}} \tag{2-2-2}$$

$\hat{X}$ 为线性近似后得到的参数估值.

(2-2-1)式中分子为 $f(X)-f(\hat{X})$的展开式中除去线性项以外的各项. 分母包括了展开式的所有余项. 显然, 若 $f(X)$是 X 的线性函数, 则分子为零, 有 $\hat{N}=0$, 否则 $\hat{N}\neq 0$. 因此, 由 $\hat{N}\neq 0$ 可以判断 $f(X)$是否为非线性函数, 且 $\hat{N}$ 越大, 非线性越强.

此定义尽管很直观, 但没有揭示"非线性"这一概念的本质. 直到 1980 年, Bates 和 Watts 受到 Efron 统计曲率定义的启示, 才从微分几何的观点出发, 定义了非线性模型的固有曲率(Intrinsic Curvature)和参数效应曲率(Parameter-effects Curvature). 他们的定义不仅反映了模型的本质, 而且计算也比(2-1-1)式简单[24](韦博成, 1989). 现介绍如下:

设非线性模型(2-1-4)式关于 X 的二阶以上连续导数存在. 现考虑参数空间 $\mathscr{X}$ 中过 X_0, 并以 h 为方向的一条直线 l_h 为[24](韦博成, 1989):

$$X(b)=X_0+bh$$

其中: b 为实参数; h 为固定方向. l 通过 $\eta=f(X)$映射到样本空间 $\mathbf{R}^n$ 中解轨迹 π 上的一条曲线 C_h 为

$$\eta=\eta_h(b)=f(X_0+bh)$$

这条曲线称为提升线(Lifted Line), 见图 2-2-1.

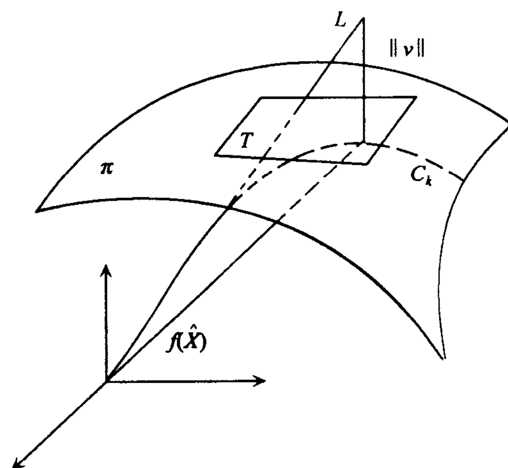


图 2-2-1

由于

$$\left.\begin{aligned}\frac{\mathrm{d}\eta_k}{\mathrm{d}b}&=\sum_{i=1}^{t}\frac{\partial f_k}{\partial x_i}\frac{\mathrm{d}x_i}{\mathrm{d}b}\\ \frac{\mathrm{d}^2\eta_k}{\mathrm{d}b^2}&=\sum_{i=1}^{t}\sum_{j=1}^{t}\frac{\partial^2 f_k}{\partial x_i\partial x_j}\frac{\mathrm{d}x_i}{\mathrm{d}b}\frac{\mathrm{d}x_j}{\mathrm{d}b}\end{aligned}\right\}\tag{2-2-3}$$

令

$$\begin{aligned}\underset{t\times 1}{h}&=\left(\frac{\mathrm{d}x_1}{\mathrm{d}b}\quad\frac{\mathrm{d}x_2}{\mathrm{d}b}\quad\cdots\quad\frac{\mathrm{d}x_t}{\mathrm{d}b}\right)'\\&=(h_1\quad h_2\quad\cdots\quad h_t)'\end{aligned}$$

$B=(b_{ki})$为 $n\times t$ 的矩阵，由(2-2-2)式定义，$W=(w_{kij})$为 $n\times t\times t$ 的立体阵(Array)(关于立体阵的有关知识详见附录 A). 由(2-2-2)式所定义的导函数矩阵再对 X 求偏导数而得，该立体阵的第 k 层由下式定义：

$$W_k=\begin{pmatrix}\frac{\partial^2 f_k}{\partial x_1^2}&\frac{\partial^2 f_k}{\partial x_1\partial x_2}&\cdots&\frac{\partial^2 f_x}{\partial x_1\partial x_t}\\ \frac{\partial^2 f_k}{\partial x_2\partial x_1}&\frac{\partial^2 f_k}{\partial x_2^2}&\cdots&\frac{\partial^2 f}{\partial x_2\partial x_t}\\ \vdots&\vdots&&\vdots\\ \frac{\partial^2 f_k}{\partial x_t\partial x_1}&\frac{\partial^2 f_k}{\partial x_t\partial x_2}&\cdots&\frac{\partial^2 f_k}{\partial x_t^2}\end{pmatrix}_{X=\hat{X}}\tag{2-2-4}$$

于是(2-2-3)式可写成

$$\left.\begin{aligned}\frac{\mathrm{d}\eta_k}{\mathrm{d}b}&=\sum_{i=1}^{t}b_{ki}h_i\\ \frac{\mathrm{d}^2\eta_k}{\mathrm{d}b^2}&=\sum_{i=1}^{t}\sum_{j=1}^{t}w_{kij}h_ih_j\\ k&=1,2,\cdots,n\end{aligned}\right\}\tag{2-2-5}$$

提升线 C_h 沿 h 方向的前二阶导数 $\dot{\eta}_h=\frac{\partial\eta}{\partial b}$, $\ddot{\eta}_h=\frac{\partial^2\eta}{\partial b^2}$, 分别为：

$$\dot{\eta}_h=Bh,\quad \ddot{\eta}_h=h'Wh\tag{2-2-6}$$

当 $b=0$ 时，(2-2-6)式对应于点 X_0，式中 η_h 对 b 的二阶导数 $\ddot{\eta}_h$ 称为加速度向量，它可分解为三个分量，即垂直于切平面的法分量 $\ddot{\eta}_h^N$，在切平面上平行于切线方向 $\dot{\eta}_h$ 和垂直于切线方向 $\dot{\eta}_h$ 的两个分量 $\ddot{\eta}_h^P$ 和 $\ddot{\eta}_h^G$. 切平面的两个分量 $\ddot{\eta}_h^P$ 和 $\ddot{\eta}_h^G$ 构成切分量 $\ddot{\eta}_h^T$，故有

$$\ddot{\eta}_h=\ddot{\eta}_h^N+\ddot{\eta}_h^P+\ddot{\eta}_h^G=\ddot{\eta}_h^N+\ddot{\eta}_h^T\tag{2-2-7}$$

(2-2-7)式中分量 $\ddot{\eta}_h^N$ 是由于解轨迹沿法方向弯曲而引起的；分量 $\ddot{\eta}_h^T$ 则是由于切平面上沿着 h 方向及其垂直方向的不均匀性而引起的. 于是非线性模型(2-1-4)式沿 h 方向在 X_0 处的固有曲率 K_h^N 和参数效应曲率 K_h^T 分别定义为

$$\left.\begin{aligned}K_h^N&=\frac{\|\ddot{\eta}_h^N\|}{\|\dot{\eta}_h\|^2}=\frac{\|(h'Wh)^N\|}{h'B'Bh}\\ K_h^T&=\frac{\|\ddot{\eta}_h^T\|}{\|\dot{\eta}_h\|^2}=\frac{\|(h'Wh)^T\|}{h'B'Bh}\end{aligned}\right\}\tag{2-2-8}$$

式中$(h'Wh)^N$和$(h'Wh)^T$分别为向量$h'Wh$的法分量和切分量. 在X_0处所有可能方向中最大的那个固有曲率和最大的那个参数效应曲率定义为非线性模型(2-1-4)的最大固有曲率和最大参数效应曲率, 即

$$\left.\begin{aligned} K^N &= \max_h K_h^N \\ K^T &= \max_h K_h^T \end{aligned}\right\} \tag{2-2-9}$$

固有曲率K_h^N是一个不依赖于坐标选择的不变量. 由于它与参数的选择无关, 只取决于模型本身的固有性质, 因此称它为固有曲率. 而K_h^T则与参数的选择有关, 即K_h^T不仅由模型本身决定, 而且还强烈地依赖于参数的选择, 所以称其为参数效应曲率[24](韦博成, 1989).

在固有曲率和参数效应曲率的定义式(2-2-8)中, h表示方向向量. 计算时可以取单位向量或其他方便的形式.

二、h方向上固有曲率和参数效应曲率的计算

若直接按定义式(2-2-8)式计算$\boldsymbol{h}$方向上的固有曲率和参数效应曲率, 需要涉及立体阵的运算. 尽管附录A中(A-5)式提供了向量$h'Wh$的计算公式, 然而公式比较复杂, 事先必须先计算出立体阵W. 立体阵W不仅书写困难, 而且立体阵的运算容易出错. 因此, 有必要在实际计算中设法避免立体阵的运算.

为此, 我们将$f(X)$对X的二阶偏导数重新排列, 将其排为一个$n\times\alpha$的矩阵C; 而将$dx_1^2,\cdots,dx_t^2,dx_1dx_2,\cdots,dx_{t-1}dx_t$用一个$\alpha$维向量$h^2$表示. 即令

$$\begin{aligned} h^2 = dX^2 &= (dx_1^2, dx_2^2, \cdots, dx_t^2, dx_1dx_2, \cdots dx_2dx_t, dx_2dx_3, \cdots, dx_{t-1}dx_t)' \\ &= (h_1^2, h_2^2, \cdots, h_t^2, h_1h_2, \cdots, h_1h_t, h_2h_3, \cdots, h_{t-1}h_t)' \end{aligned} \tag{2-2-10}$$

$$C = \begin{bmatrix} \frac{\partial^2 f_1}{\partial x_1^2} & \frac{\partial^2 f_1}{\partial x_2^2} & \cdots & \frac{\partial^2 f_1}{\partial x_t^2} & 2\frac{\partial^2 f_1}{\partial x_1\partial x_2} & \cdots & 2\frac{\partial^2 f_1}{\partial x_1\partial x_t} & 2\frac{\partial^2 f_1}{\partial x_2\partial x_3} & \cdots & 2\frac{\partial^2 f_1}{\partial x_{t-1}\partial x_t} \\ \frac{\partial^2 f_2}{\partial x_1^2} & \frac{\partial^2 f_2}{\partial x_2^2} & \cdots & \frac{\partial^2 f_2}{\partial x_t^2} & 2\frac{\partial^2 f_2}{\partial x_1\partial x_2} & \cdots & 2\frac{\partial^2 f_2}{\partial x_1\partial x_t} & 2\frac{\partial^2 f_2}{\partial x_2\partial x_3} & \cdots & 2\frac{\partial^2 f_2}{\partial x_{t-1}\partial x_t} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial^2 f_n}{\partial x_1^2} & \frac{\partial^2 f_n}{\partial x_2^2} & \cdots & \frac{\partial^2 f_n}{\partial x_t^2} & 2\frac{\partial^2 f_n}{\partial x_1\partial x_2} & \cdots & 2\frac{\partial^2 f_n}{\partial x_1\partial x_t} & 2\frac{\partial^2 f_n}{\partial x_2\partial x_3} & \cdots & 2\frac{\partial^2 f_n}{\partial x_{t-1}\partial x_t} \end{bmatrix}_{X=X_0} \tag{2-2-11}$$

式中:

$$\alpha = \frac{1}{2}t(t+1) \tag{2-2-12}$$

于是下列等式成立:

$$\ddot{\eta}_h = h'Wh = Ch^2 \tag{2-2-13}$$

证明: 设

$$\beta = h'Wh, \gamma = Ch^2$$

由附录A中(A-4)式知

$$\beta = \left(\sum_{i=1}^{t}\sum_{j=1}^{t} W_{1ij}h_ih_j \sum_{i=1}^{t}\sum_{j=1}^{t} W_{2ij}h_ih_j \cdots \sum_{i=1}^{t}\sum_{j=1}^{t} W_{nij}h_ih_j\right)'$$

则

$$\beta_k = \sum_{i=1}^{t}\sum_{j=1}^{t} W_{kij}h_ih_j$$

顾及(2-2-4)式,得

$$\beta_k = \frac{\partial^2 f_k}{\partial x_1^2}h_1^2 + \frac{\partial^2 f_k}{\partial x_1 \partial x_2}h_1h_2 + \cdots + \frac{\partial^2 f_k}{\partial x_1 \partial x_t}h_1h_t + \frac{\partial^2 f_k}{\partial x_2 \partial x_1}h_2h_t + \frac{\partial^2 f_k}{\partial x_2^2}h_2^2 + \cdots$$

$$+ \frac{\partial^2 f_k}{\partial x_2 \partial x_t}h_2h_t + \cdots + \frac{\partial^2 f_k}{\partial x_t \partial x_1}h_th_1 + \frac{\partial^2 f_k}{\partial x_t \partial x_2}h_th_2 + \cdots + \frac{\partial^2 f_k}{\partial x_t^2}h_t^2$$

因为$\frac{\partial^2 f_k}{\partial x_i \partial x_j}h_ih_j = \frac{\partial^2 f_k}{\partial x_j \partial x_i}h_jh_i$,所以

$$\beta_k = \sum_{i=1}^{t}\frac{\partial^2 f_k}{\partial x_i^2}h_i^2 + 2\sum_{i=1}^{t-1}\sum_{j=i+1}^{t}\frac{\partial^2 f_k}{\partial x_i \partial x_j}h_ih_j$$

而根据(2-2-10)式和(2-2-11)式有

$$\gamma_k = \sum_{i=1}^{t}\frac{\partial^2 f_k}{\partial x_i^2}h_i^2 + 2\sum_{i=1}^{t-1}\sum_{j=1}^{t}\frac{\partial^2 f_k}{\partial x_i \partial x_j}h_ih_j = \beta_k$$

所以(2-2-13)式成立.

于是(2-2-8)式可写为

$$\left.\begin{aligned} K_h^N &= \frac{\|(Ch^2)^N\|}{h'B'Bh} = \frac{\|\ddot{\eta}_h^N\|}{h'B'Bh} \\ K_h^T &= \frac{\|(Ch^2)^T\|}{h'B'Bh} = \frac{\|\ddot{\eta}_h^T\|}{h'B'Bh} \end{aligned}\right\} \tag{2-2-14}$$

根据(2-2-13)式,可以将提升线 C_h 沿h 方向的二阶导数$\ddot{\eta}_h$ 表示成一个$n\times\alpha$ 的矩阵与一个 α 维向量的乘积.有了 $\ddot{\eta}_h$ 后,还必须计算 $\ddot{\eta}_h$ 的法分量$\ddot{\eta}_h^N$ 和切分量$\ddot{\eta}_h^T$.为此,将由 B 生成的投影矩阵记为 S_T,并记 $S_N = I - S_T$,于是有

$$\left.\begin{aligned} \ddot{\eta}_h^T &= S_T\ddot{\eta}_h \\ \ddot{\eta}_h^N &= S_N\ddot{\eta}_h = \ddot{\eta}_h - \ddot{\eta}_h^T \end{aligned}\right\} \tag{2-2-15}$$

式中:

$$\left.\begin{aligned} S_T &= B(B'B)^{-1}B' \\ S_N &= I - S_T \end{aligned}\right\} \tag{2-2-16}$$

有了(2-2-13)式、(2-2-14)式和(2-2-15)式后,就很容易按(2-2-8)式计算非线性模型 h 方向上的固有曲率K_h^N 和参数效应曲率K_h^T 了.其计算步骤如下:

(1) 适当选取参数的近似值 X_0,在 X_0 处将非线性函数展为泰勒级数,取至二次项.并用 X_0 按(2-2-2)式和(2-2-11)式计算一、二阶偏导数阵 B 和C.

(2) 适当选取 h,并按(2-2-6)式的第一式和(2-2-13)式计算沿 h 方向的前二阶导数$\dot{\eta}_h$ 和 $\ddot{\eta}_h$.

(3) 按(2-2-16)式的第一式计算投影矩阵 S_T.

(4) 按(2-2-15)式计算 $\ddot{\eta}_h$ 的切分量$\ddot{\eta}_h^T$ 和法分量$\ddot{\eta}_h^N$.

(5) 按(2-2-14)式计算沿 h 方向的固有曲率K_h^N 和参数效应曲率K_h^T.

例2-2-1 计算例2-1-1中的非线性模型沿 h 方向的固有曲率和参数效应曲率

(取 $h = \mathrm{d}X$).

由例 2-1-1 知

$$B = \begin{pmatrix} 0.7408 & 4.0004 \\ 0.5488 & 5.9272 \\ 0.4066 & 6.5864 \\ 0.3012 & 6.5058 \\ 0.2231 & 6.0245 \end{pmatrix}, \ \mathrm{d}X = \begin{pmatrix} -0.005858021 \\ 0.049953787 \end{pmatrix} = h$$

则 $h^2 = (h_1^2 \quad h_2^2 \quad h_1 h_2)' = (0.000034319 \quad 0.002495381 \quad -0.000292630)'$

$$C = \begin{pmatrix} \frac{\partial^2 f_1}{\partial x_1^2} & \frac{\partial^2 f_1}{\partial x_2^2} & 2\frac{\partial^2 f_1}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f_2}{\partial x_1^2} & \frac{\partial^2 f_2}{\partial x_2^2} & 2\frac{\partial^2 f_2}{\partial x_1 \partial x_2} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 f_5}{\partial x_1^2} & \frac{\partial^2 f_5}{\partial x_2^2} & 2\frac{\partial^2 f_5}{\partial x_1 \partial x_2} \end{pmatrix}_{X = X_0} = \begin{pmatrix} 0 & 4.0004 & 1.4816 \\ 0 & 11.8543 & 2.1952 \\ 0 & 19.7593 & 2.4394 \\ 0 & 26.0232 & 2.4096 \\ 0 & 30.1226 & 2.2313 \end{pmatrix}$$

$$\dot{\eta}_h = Bh = \begin{pmatrix} 0.7408 & 4.0004 \\ 0.5488 & 5.9272 \\ 0.4066 & 6.5864 \\ 0.3012 & 6.5058 \\ 0.2231 & 6.0245 \end{pmatrix} \begin{pmatrix} -0.005858021 \\ 0.049953787 \end{pmatrix} = \begin{pmatrix} 0.1955 \\ 0.2929 \\ 0.3266 \\ 0.3232 \\ 0.2996 \end{pmatrix}$$

$$\ddot{\eta}_h = Ch^2 = \begin{pmatrix} 0 & 4.0004 & 1.4816 \\ 0 & 11.8543 & 2.1952 \\ 0 & 19.7593 & 2.4394 \\ 0 & 26.0232 & 2.4096 \\ 0 & 30.1226 & 2.2313 \end{pmatrix} \begin{pmatrix} 0.000034 \\ 0.002495 \\ -0.000293 \end{pmatrix} = \begin{pmatrix} 0.0095 \\ 0.0289 \\ 0.0486 \\ 0.0642 \\ 0.0745 \end{pmatrix}$$

由例 2-1-1 知

$$(B'B)^{-1} = \begin{pmatrix} 3.3737 & -0.2377 \\ -0.2377 & 0.0225 \end{pmatrix}$$

故

$$S_T = B(B'B)^{-1}B' = \begin{pmatrix} 0.8030 & 0.3401 & 0.0632 & -0.0931 & -0.1726 \\ 0.3401 & 0.2610 & 0.2000 & 0.1531 & 0.1171 \\ 0.0632 & 0.2000 & 0.2617 & 0.2779 & 0.2682 \\ -0.0931 & 0.1531 & 0.2779 & 0.3278 & 0.3331 \\ -0.1726 & 0.1171 & 0.2682 & 0.3331 & 0.3464 \end{pmatrix}$$

于是

$$\ddot{\eta}_h^T = \begin{pmatrix} 0.8030 & 0.3401 & 0.0632 & -0.0931 & -0.1726 \\ 0.3401 & 0.2610 & 0.2000 & 0.1531 & 0.1171 \\ 0.0632 & 0.2000 & 0.2617 & 0.2779 & 0.2682 \\ -0.0931 & 0.1531 & 0.2779 & 0.3278 & 0.3331 \\ -0.1726 & 0.1171 & 0.2682 & 0.3331 & 0.3464 \end{pmatrix} \begin{pmatrix} 0.0095 \\ 0.0289 \\ 0.0486 \\ 0.0642 \\ 0.0745 \end{pmatrix} = \begin{pmatrix} 0.0017 \\ 0.0390 \\ 0.0569 \\ 0.0629 \\ 0.0620 \end{pmatrix}$$

$$\ddot{\eta}_h^N = \ddot{\eta}_h - \ddot{\eta}_h^T = \begin{pmatrix} 0.0078 \\ -0.0101 \\ -0.0083 \\ 0.0013 \\ 0.0125 \end{pmatrix}$$

所以

$$K_h^N = \frac{\sqrt{0.000390}}{0.4249} = 0.0465, \quad K_h^T = \frac{\sqrt{0.012855}}{0.4249} = 0.2638$$

例 2-2-2 计算例2-1-2中的非线性模型沿 h 方向的固有曲率和参数效应曲率.

由例 2-1-2 知：$B = \begin{pmatrix} 10.8 & 1 \\ 5.4 & 1 \\ 3.6 & 1 \\ 2.7 & 1 \\ 2.16 & 1 \end{pmatrix}$，取 $h = \begin{pmatrix} -0.005858021 \\ 0.049953787 \end{pmatrix}$

则 $\quad h^2 = (h_1^2 \quad h_2^2 \quad h_1 h_2)' = (0.000034319 \quad 0.002495381 \quad -0.000292630)'$

$$C = \begin{pmatrix} \frac{\partial^2 f_1}{\partial x_1^2} & \frac{\partial^2 f_1}{\partial x_2^2} & 2\frac{\partial^2 f_1}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f_2}{\partial x_1^2} & \frac{\partial^2 f_2}{\partial x_2^2} & 2\frac{\partial^2 f_2}{\partial x_1 \partial x_2} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 f_5}{\partial x_1^2} & \frac{\partial^2 f_5}{\partial x_2^2} & 2\frac{\partial^2 f_5}{\partial x_1 \partial x_2} \end{pmatrix}_{X=X_0} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 0 & 0 \\ 0.6667 & 0 & 0 \\ 0.5 & 0 & 0 \\ 0.4 & 0 & 0 \end{pmatrix}$$

$$\dot{\eta}_h = Bh = \begin{pmatrix} 10.8 & 1 \\ 5.4 & 1 \\ 3.6 & 1 \\ 2.7 & 1 \\ 2.16 & 1 \end{pmatrix} \begin{pmatrix} -0.005858021 \\ 0.049953787 \end{pmatrix} = \begin{pmatrix} -0.0133 \\ 0.0183 \\ 0.0289 \\ 0.0341 \\ 0.0373 \end{pmatrix}$$

$$\ddot{\eta}_h = Ch^2 = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 0 & 0 \\ 0.6667 & 0 & 0 \\ 0.5 & 0 & 0 \\ 0.4 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.000034319 \\ 0.002495381 \\ -0.000292630 \end{pmatrix} = \begin{pmatrix} 0.0000686 \\ 0.0000343 \\ 0.0000229 \\ 0.0000172 \\ 0.0000137 \end{pmatrix}$$

由例 2-1-2 知

$$(B'B)^{-1}=\begin{pmatrix}0.0204 & -0.1005\\ -0.1005 & 0.6955\end{pmatrix}$$

故

$$S_T=B(B'B)^{-1}B'=\begin{pmatrix}0.9042 & 0.2571 & 0.0415 & -0.0664 & -0.1311\\ 0.2571 & 0.2050 & 0.1876 & 0.1789 & 0.1737\\ 0.0415 & 0.1876 & 0.2363 & 0.2606 & 0.2753\\ -0.0664 & 0.1789 & 0.2606 & 0.3015 & 0.3260\\ -0.1311 & 0.1737 & 0.2753 & 0.3260 & 0.3556\end{pmatrix}$$

于是

$$\ddot{\eta}_h^T=\begin{pmatrix}0.9042 & 0.2571 & 0.0415 & -0.0664 & -0.1311\\ 0.2571 & 0.2050 & 0.1876 & 0.1789 & 0.1738\\ 0.0415 & 0.1876 & 0.2363 & 0.2606 & 0.2753\\ -0.0664 & 0.1789 & 0.2606 & 0.3015 & 0.3260\\ -0.1311 & 0.1737 & 0.2753 & 0.3260 & 0.3556\end{pmatrix}\begin{pmatrix}0.0000686\\ 0.0000343\\ 0.0000229\\ 0.0000172\\ 0.0000137\end{pmatrix}=\begin{pmatrix}0.0000689\\ 0.0000344\\ 0.0000229\\ 0.0000172\\ 0.0000137\end{pmatrix}$$

$$\ddot{\eta}_h^N=\begin{pmatrix}-1.0\times10^{-7}\\ -1.0\times10^{-7}\\ 0\\ 0\\ 0\end{pmatrix}$$

所以

$$K_h^N=\frac{\sqrt{2\times10^{-14}}}{0.0039}=0.000036,\quad K_h^T=\frac{\sqrt{6.8943\times10^{-9}}}{0.0039}=0.0213$$

由计算知，例 2-1-2 中的非线性模型的固有曲率和参数效应曲率都远远小于例 2-1-1 中的非线性模型的固有曲率和参数效应曲率. 所以例 2-1-2 中线性近似所引起的模型误差小于例 2-1-1 中线性近似所产生的模型误差.

第三节　曲率立体阵

前节定义的非线性模型的固有曲率和参数效应曲率，都与方向向量 h 有关. 能否引进一些量，使它既能刻画非线性模型的本质，又与方向向量 h 无关呢？本节要介绍的曲率立体阵就是如此.

一、曲率立体阵的定义

当我们用参数的近似值 X_0 计算矩阵 B 时，B 矩阵的列向量就是 X_0 处切空间的一组基. 现对这组基进行变换，将其变换为一组标准正交基. 为此，可对 B 的列向量进行正交化，即对矩阵 B 进行 QR 分解[24]（韦博成，1989）：

$$B=(Q,N)\begin{pmatrix}R\\ 0\end{pmatrix}=QR \tag{2-3-1}$$

其中 Q 和 N 的列向量分别为 X_0 处切空间的一组标准正交基和法空间的一组标准正交基.

R 为非退化上三角阵.

这个正交化过程相当于对参数 X 进行了一个变换:

$$\varphi = R(X - X_0),\ X = X_0 + M\varphi,\ M = R^{-1} \tag{2-3-2}$$

Q 的列向量就是切空间关于新坐标 φ 的一组标准正交基. 这是因为此时 $\eta = \eta(\varphi) = f(X_0 + M\varphi)$, 且

$$\frac{\partial \eta_k}{\partial \varphi_j} = \sum_{i=1}^{t} \frac{\partial \eta_k}{\partial X_i} \frac{\partial X_i}{\partial \varphi_j} \quad (j = 1, 2, \cdots, t; k = 1, 2, \cdots, n)$$

即

$$\frac{\partial \eta}{\partial \varphi} = \frac{\partial \eta}{\partial X} \frac{\partial X}{\partial \varphi} = BM = Q \tag{2-3-3}$$

现在再在 φ 坐标和标准正交基下来推导固有曲率和参数效应曲率. 为此, 求 $\eta = \eta(\varphi)$ 关于 φ 的二阶导数, 由(2-3-2)式知, φ 是 X 的线性函数, 故有

$$\frac{\partial^2 \eta_k}{\partial \varphi_r \partial \varphi_s} = \sum_{i=1}^{t} \sum_{j=1}^{t} \frac{\partial \eta_k}{\partial x_i \partial x_j} \frac{\partial x_i}{\partial \varphi_r} \frac{\partial x_j}{\partial \varphi_s}$$

式中: $r, s = 1, 2, \cdots, t; k = 1, 2, \cdots, n$.

令

$$U = \frac{\partial^2 \eta}{\partial \varphi^2} = M'WM \tag{2-3-4}$$

式中: U 为 $n \times t \times t$ 阶立体阵, 即

$$U = (u_{kij}), \quad u_{kij} = \frac{\partial^2 \eta_k}{\partial \varphi_i \partial \varphi_j} \tag{2-3-5}$$

经过(2-3-2)式的变换, 参数空间中过 X_0 以 h 为方向的直线 $X = X_0 + bh$ 变为

$$\varphi = R(X - X_0) = R(X_0 + bh - X_0) = bRh = bd$$

式中: $d = Rh$ 为方向向量. 而提升线变为 $\eta_d = f(bd)$, 且对应于 $\varphi = 0$. 可见, 通过变换, (2-2-8)式中的 B, W, h 分别变为 Q, U 和 d. 于是取 d 为单位向量, 将(2-3-3)式和(2-3-4)式代入(2-2-8)式, 得 φ 坐标下沿 d 方向的曲率为

$$K_d^N = \| (d'Ud)^N \|,\ K_d^T = \| (d'Ud)^T \| \tag{2-3-6}$$

由(2-3-3)式和(2-3-1)式知, Q 和 N 的列向量分别生成解轨迹在 $\varphi = 0$ 处的切空间和法空间. 因此, 相应的投影矩阵 S_T 和 S_N 可分别表示为

$$S_T = QQ',\ S_N = NN' \tag{2-3-7}$$

于是有

$$(d'Ud)^T = (QQ')(d'Ud)$$

利用矩阵与立阵的方括号乘积, 即附录中的(A-8)式及其性质 7 和性质 10, 有

$$(d'Ud)^T = [QQ'][d'Ud] = [Q][[Q'][d'Ud]] = [Q][d'[Q'][U]d]$$

因为 Q 是正交矩阵, 所以有 $Q'Q = I$. 于是根据向量范数(欧氏范数)的定义可得:

$$\| (d'Ud)^T \| = \| d'[Q'][U]d \|$$

同理可得

$$\| (d'Ud)^N \| = \| d'[N'][U]d \|$$

即

$$\left.\begin{aligned} K_d^N &= \| d'[N'][U]d \| \\ K_d^T &= \| d'[N'][U]d \| \end{aligned}\right\} \tag{2-3-8}$$

根据(2-3-8)式，我们给出下列定义[24]（韦博成，1989）.

定义 2-3-1 非线性模型(2-1-4)式的固有曲率立体阵 G 和参数效应立体阵 H 分别定义为

$$\left.\begin{aligned} G &\triangleq [N'][U] \\ H &\triangleq [Q'][U] \end{aligned}\right\} \tag{2-3-9}$$

式中：G 为 $(n-t)\times t\times t$ 的立体阵；H 为 $t\times t\times t$ 的立体阵. 它们仅与模型及参数有关，而与方向 d 无关.

二、曲率立体阵的计算

由前面的推导知，计算曲率立体阵的关键是 B 矩阵的 QR 分解. 当 B 矩阵列满秩时，$B'B$ 为正定阵. 于是根据乔勒斯基(Cholesky)分解，可得：

$$B'B = T'T$$

式中：T 为上三角阵，其非零元素可由下式计算[45]（王新洲，1991）：

$$\left.\begin{aligned} t_{ii} &= \sqrt{b_{ii}} \\ t_{ij} &= \frac{b_{ij}}{t_{ii}} \end{aligned}\right\} \quad (j > i = 1) \tag{2-3-10}$$

$$\left.\begin{aligned} t_{ii} &= \sqrt{b_{ii} - \sum_{k=1}^{i-1} t_{ki}^2} \\ t_{ij} &= \frac{\left(b_{ij} - \sum_{k=1}^{i-1} t_{ki}t_{kj}\right)}{t_{ii}} \end{aligned}\right\} \quad (j > i > 1), \tag{2-3-11}$$

式中：b_{ij} 为正定阵 $B'B$ 的第 i 行第 j 列元素，即 $B'B=(b_{ij})$.

求得上三角阵 T 后，由下式可得 B 的 QR 分解：

$$\left.\begin{aligned} B &= (BT^{-1})T \\ Q &= BT^{-1} \\ R &= T \end{aligned}\right\} \tag{2-3-12}$$

至于 N 矩阵，可根据正交条件求出.

例 2-3-1 设某非线性模型的矩阵 B 和立阵 W 分别为：$B=\begin{pmatrix}1\\-2\\2\end{pmatrix}$，$W=\begin{bmatrix}3\\2\\1\end{bmatrix}$，求其曲率立体阵.

因为 $n=3$，$t=1$，所以有 $B'B=9$.

由(2-3-10)式得，$T=3$，所以 $T^{-1}=\frac{1}{3}$. 于是 $Q=BT^{-1}=\frac{1}{3}\begin{pmatrix}1\\-2\\2\end{pmatrix}$，$R=3$，$M=T^{-1}$

$=\frac{1}{3}$.

设 $N=\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \\ n_{31} & n_{32} \end{pmatrix}$,由正交条件 $Q'N=0, N'N=I$ 可得:

$$\begin{cases} n_{11}-2n_{21}+2n_{31}=0 \\ n_{12}-2n_{22}+2n_{32}=0 \\ n_{11}^2+n_{21}^2+n_{31}^2=1 \\ n_{12}^2+n_{22}^2+n_{32}^2=1 \\ n_{11}n_{12}+n_{21}n_{22}+n_{31}n_{32}=0 \end{cases} \tag{2-3-13}$$

令 $n_{11}=0.1$,由(2-3-13)式的第一式和第三式可解得 $n_{21}=0.728118$, $n_{31}=0.678118$,再由(2-3-13)式的第二式、第四式和第五式可解得

$$n_{12}=0.937491, \quad n_{22}=0.159373 \quad n_{32}=-0.309373$$

于是

$$N=\begin{pmatrix} 0.1 & 0.937491 \\ 0.728118 & 0.159373 \\ 0.678118 & -0.309373 \end{pmatrix}, \quad U=M'WM=\frac{1}{3}\begin{pmatrix}3\\2\\1\end{pmatrix}\frac{1}{3}=\frac{1}{9}\begin{pmatrix}3\\2\\1\end{pmatrix}$$

所以由定义(2-3-9)式得

$$G=\left[\begin{pmatrix} 0.1 & 0.937491 \\ 0.728118 & 0.159373 \\ 0.678118 & -0.309373 \end{pmatrix}'\right]\left[\frac{1}{9}\begin{pmatrix}3\\2\\1\end{pmatrix}\right]=\frac{1}{9}\begin{pmatrix}2.434354\\2.821846\end{pmatrix}$$

$$H=\left[\frac{1}{3}\begin{pmatrix}1\\-2\\2\end{pmatrix}'\right]\left[\frac{1}{9}\begin{pmatrix}3\\2\\1\end{pmatrix}\right]=\frac{1}{27}$$

显然,当观测值的个数 n 和未知参数的个数 t 较大时,曲率立体阵的计算很复杂.这时可采用专门的计算机程序计算.

第四节　非线性模型线性近似的容许曲率

由于不同非线性模型的固有曲率和参数效应曲率不同,有的较小,即非线性强度较弱;有的较大,即非线性强度较强.因此,有的非线性模型线性近似时只产生较小的模型误差,得到令人满意的结果.而有的非线性模型对参数的近似值十分敏感,线性近似时会产生较大的模型误差.还有些非线性模型甚至不能线性近似.既然有的非线性模型可以线性近似,有的则不行,那么,如何判断一个非线性模型可否线性近似呢?下面就来讨论这个问题.

一、非线性模型线性近似的容许曲率[39](王新洲,1997)

将非线性模型(2-1-4)式在 X_0 处线性近似,得

$$L\approx f(X_0)+B(X-X_0)+\Delta \tag{2-4-1}$$

相应的误差方程为：

$$V = B\mathrm{d}X + l \tag{2-4-2}$$

式中：

$$\mathrm{d}X = \hat{X} - X_0,\ l = f(X_0) - L$$

在假定 Δ 相互独立且服从 $N(0, \sigma^2 Q_{LL})$ 的前提下：

$$\| f(X) - f(\hat{X}) \|^2 \approx (X - \hat{X}) B'(\hat{X}) PB(\hat{X})(X - \hat{X})$$

于是，统计量

$$K_1 = \frac{(X - \hat{X}) B'(\hat{X}) PB(\hat{X})(X - \hat{X})}{\sigma^2} \sim \chi^2(t)$$

因为[46]（李庆海，陶本藻，1982）：

$$K_2 = \frac{V'V}{\sigma^2} = \frac{(n-t)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-t)$$

于是

$$\frac{\dfrac{K_1}{t}}{\dfrac{K_2}{n-t}} = \frac{\dfrac{\| f(X) - f(\hat{X}) \|^2}{\sigma^2 t}}{\dfrac{\hat{\sigma}^2(n-t)}{\sigma^2(n-t)}} = \frac{\| f(X) - f(\hat{X}) \|^2}{t\hat{\sigma}^2} \sim F(t, n-t) \tag{2-4-3}$$

所以(2-4-1)式在置信水平$(1-\alpha)$下的置信域为：

$$\begin{aligned} \| f(X) - f(\hat{X}) \|^2 &\approx (X - \hat{X})' B'(\hat{X}) PB(\hat{X})(X - \hat{X}) \\ &\leqslant t\hat{\sigma}^2 F(t, n-t) \end{aligned} \tag{2-4-4}$$

式中：t 为未知参数 X 的个数；$\hat{\sigma}^2 = \dfrac{V'V}{n-t}$ 为方差 σ^2 的估值；$B(\hat{X})$ 表示用线性近似时的最小二乘估值 $\hat{X}$ 计算出来的 B 矩阵的值.

置信域(2-4-4)的边界可以看成一个以 $f(\hat{X})$ 为球心，以 $R = \rho\sqrt{F}$ 为半径的球面. 其中 $\rho = \hat{\sigma}\sqrt{t}$，显然，这个球面上任一点的曲率为

$$K_F = \frac{1}{\rho\sqrt{F}} \tag{2-4-5}$$

(2-4-5)式是线性近似后，在置信水平 $1-\alpha$ 下置信域的曲率. 当非线性模型(2-1-4)式的解轨迹 $\eta = f(X)$ 上 X_0 处的最大固有曲率 K^N 和最大参数效应曲率 K^T 均小于 K_F 时，说明解轨迹接近于线性. 即当

$$\left.\begin{aligned} K^N < K_F = \frac{1}{\rho\sqrt{F}} \\ K^T < K_F = \frac{1}{\rho\sqrt{F}} \end{aligned}\right\} \tag{2-4-6}$$

成立时，就认为解轨迹接近线性.

令

$$\gamma_h^N = \rho K_h^N, \quad \gamma_h^T = \rho K_h^T \tag{2-4-7}$$

并称 γ_h^N 和 γ_h^T 分别为相对固有曲率和相对参数效应曲率. 简称相对曲率. 仿(2-4-7)式，可定义最大相对曲率为

$$\Gamma^N = K^N,\ \Gamma^T = K^T \tag{2-4-8}$$

且当

$$\Gamma^N < \frac{1}{\sqrt{F}},\ \Gamma^T < \frac{1}{\sqrt{F}} \tag{2-4-9}$$

成立时,非线性模型(2-1-4)式的解轨迹就接近线性,于是非线性模型(2-1-4)式就可以线性近似,否则将不能线性近似.因此,可定义非线性模型线性近似的容许曲率.

定义 2-4-1 非线性模型(2-1-4)式线性近似的容许曲率定义为

$$\Gamma_{容} = \frac{1}{\sqrt{F}} \tag{2-4-10}$$

容许曲率 $\Gamma_{容}$ 也是一个相对曲率.

二、判断非线性模型能否线性近似的方法

定义了非线性模型线性近似的容许曲率后,就能对任一非线性模型按如下方法判断能否线性近似.

(1)首先对具体的非线性模型计算最大相对固有曲率 Γ^N 和最大相对参数效应曲率 Γ^T,Γ^N 和 Γ^T 的计算一般比较复杂,需用专门的计算机程序在计算机上计算.

(2)根据分子自由度 t,分母自由度 $n-t$ 和显著水平 α 查 F 分布表,得临界值 F.并按(2-4-10)式计算容许曲率 $\Gamma_{容}$.

(3)检查(2-4-9)式是否成立.若(2-4-9)式成立,即 Γ^N 和 Γ^T 都小于 $\Gamma_{容}$,则该非线性模型可以线性近似.如果(2-4-9)式不成立,这时可能出现三种情况:第一种情况,Γ^N 和 Γ^T 都大于 $\Gamma_{容}$,说明非线性模型的非线性强度很强,不能线性近似.第二种情况,$\Gamma^N > \Gamma_{容}$,而 $\Gamma^T < \Gamma_{容}$.说明非线性模型的固有曲率很大,只是参数选择较好.由于固有曲率很大,表明非线性模型的非线性强度很强,所以仍不能线性近似.第三种情况,$\Gamma^N < \Gamma_{容}$,而 $\Gamma^T > \Gamma_{容}$.表明非线性模型的非线性强度较弱,只要设法对参数进行变换,使得在新参数下的参数效应曲率小于容许曲率,则在新参数下可以对模型进行线性近似.

例 2-4-1 计算例2-1-1和例 2-1-2 中两非线性模型的容许曲率.

因为 $t=2, n=5$,取 $\alpha=0.05$,则由 F 分布表[46](李庆海,陶本藻,1982),可得 $F(2,3,0.05)=9.6$.所以此两非线性模型的容许曲率为

$$\Gamma_{容} = \frac{1}{\sqrt{F}} = 0.322749$$

由例2-1-1 知,$\hat{\sigma} = \sqrt{\frac{V'V}{n-t}} = \pm 0.0418$,故 $\rho = \hat{\sigma}\sqrt{t} = 0.0418\sqrt{2} = 0.0591$,则沿 h 方向的相对固有曲率和相对参数效应曲率分别为

$$\gamma_h^N = 0.0643 \times 0.0591 = 0.0038$$

$$\gamma_h^T = 0.2844 \times 0.0591 = 0.0168$$

有

$$\gamma_h^N < \Gamma_{容}, \quad \gamma_h^T < \Gamma_{容}$$

同样由例 2-1-1 知,$\hat{\sigma} = \sqrt{\frac{V'V}{n-t}} = \pm 0.006567$,故 $\rho = 0.006567\sqrt{2} = 0.009287$,于是沿 h 方向的相对固有曲率和相对参数效应曲率分别为

$$\gamma_h^N = 0.00003 \times 0.009281 = 2.7861 \times 10^{-7} < \Gamma_{容}$$

$$\gamma_h^T = 0.008697 \times 0.009287 = 8.0769 \times 10^{-5} < \Gamma_{容}$$

此两例的相对固有曲率和相对参数效应曲率都小于容许曲率，但一般不能据此作出能线性近似的结论. 这是因为(2-4-9)式中用的是非线性模型的最大相对固有曲率和最大相对参数效应曲率，而上面计算的却是沿一特定方向 h 上的相对固有曲率和相对参数效应曲率. 要根据(2-4-9)式判断一个非线性模型能否线性近似，必须计算该非线性模型的最大相对曲率. 而最大相对曲率计算较困难，所以(2-4-9)式不便于实际应用. 为此笔者重新定义了一个方便计算的判断准则[39]（王新洲，1997）.

三、判断非线性模型能否线性近似的实用准则

将 $f(X)$ 展为泰勒级数，取至二次项得

$$f(X) = f(X_0) + B\mathrm{d}X + \frac{1}{2}C\mathrm{d}X^2 + e$$

式中：e 为略去三次及三次以上各项后引起的误差向量.

当 $f(X)$ 为线性模型时，有 $\frac{1}{2}C\mathrm{d}X^2 = e = 0$. 于是有

$$\| f(X) - f(X_0) - B\mathrm{d}X \|^2 \geqslant \| C\mathrm{d}X^2 \|^2 \tag{2-4-11}$$

用观测值 L 和线性近似后解出的 $\mathrm{d}X$ 代入(2-4-11)式，得

$$\| -V \|^2 \geqslant \| C\mathrm{d}X^2 \|^2 \tag{2-4-12}$$

或写为

$$V'V \geqslant \| C\mathrm{d}X^2 \|^2 \tag{2-4-12}$$

(2-4-12)式表明：线性近似后的残差平方和大于、等于略去二次项所产生的误差向量的平方和. 这是可以理解的，因为 $C\mathrm{d}X^2$ 只是线性近似所产生的模型误差向量，而 V 既包括这一模型误差向量，又包括观测误差的影响. 所以，如(2-4-12)式不成立，就说明略去二次项产生的模型误差大于观测误差，此时就不能线性近似. 因此，由(2-4-12)式可得到判断非线性模型能否线性近似的笼统判断准则为[41]（王新洲，1999）：

$$\frac{\| C\mathrm{d}X^2 \|^2}{V'V} \leqslant 1 \tag{2-4-13}$$

例 2-4-2 根据(2-4-13)式判断例 2-1-1 和例 2-1-2 的两个非线性模型能否线性近似.

由例 2-1-1 知

$$V = (-0.0041 \quad 0.0065 \quad 0.0021 \quad -0.0004 \quad -0.0055)'$$

又由例 2-2-1 知

$$C\mathrm{d}X^2 = (0.0095 \quad 0.0289 \quad 0.0486 \quad 0.0642 \quad 0.0745)'$$

于是

$$\frac{\| C\mathrm{d}X^2 \|^2}{V'V} = \frac{0.012959}{0.000093} = 139.3474 > 1$$

所以例 2-1-1 中的非线性模型不能线性近似.

另由例 2-1-2 知

$$V=(-0.0010 \quad 0.0010 \quad 0.0049 \quad -0.0031 \quad -0.0019)'$$

$$dX^2=(0.0004 \quad 0.0018 \quad 0.0009)'$$

而 C 已由例2-2-2算出,故

$$CdX^2=(0.0008 \quad 0.0004 \quad 0.0003 \quad 0.0002 \quad 0.0002)'$$

于是

$$\frac{\| CdX^2 \|^2}{V'V}=\frac{9.381\times 10^{-7}}{3.9\times 10^{-5}}=0.0241<1$$

所以例2-1-2中的非线性模型可以线性近似.

根据这两个结论,我们就不难理解为什么例2-1-1中参数估值 $\hat{X}$ 与其真值 X 的距离要比例2-1-2中参数估值 $\hat{X}$ 与其真值 X 的距离大一个数量级了.

四、固有非线性性和参数效应非线性性

前已述及(2-2-8)式揭示了非线性模型的本质.故用(2-4-9)式判断非线性模型能否线性近似比较全面,又能回答非线性模型不能线性近似的原因是非线性模型的固有特性所致,还是由参数选择不当所致.若仅是非线性模型的参数效应曲率较大,而固有曲率不大,则可通过参数变换,使得在新参数下的参数效应曲率很小或为零.则在新参数下可对非线性模型进行线性近似.而用(2-4-13)式判断,尽管计算比较简单,判断也较准确,但当通过判断,得出非线性模型不能线性近似时,却不能回答是此非线性模型的固有非线性性本身较强,还是仅参数效应非线性性较强.即不能回答此模型是否可通过参数变换后再线性近似.为了解决这个问题,下面仿定义固有曲率和参数效应曲率一样,来定义固有非线性性和参数效应非线性性.

将 CdX^2 投影到解轨迹在 $X=X_0$ 处的切空间和法空间,则有

$$CdX^2=C^N dX^2+C^T dX^2 \tag{2-4-14}$$

式中:$C^N=S_N C$;$C^T=S_T C$,投影矩阵 S_N、S_T 由(2-2-16)式确定.

证明:

$$C^N dX^2+C^T dX^2=(C^N+C^T)dX^2=(S_N C+S_T C)dX^2$$
$$=(S_N+S_T)CdX^2$$

由(2-2-16)式知

$$S_N+S_T=I$$

所以,$C^N dX^2+C^T dX^2=CdX^2$,(2-4-14)式得证.

由(2-4-14)式可得

$$\| CdX^2 \| = \| C^N dX^2+C^T dX^2 \|^2 \leqslant \| C^N dX^2 \|^2+\| C^T dX^2 \|^2$$

所以有

$$\left.\begin{array}{l}\| CdX^2 \|^2 \geqslant \| C^N dX^2 \|^2 \\ \| CdX^2 \|^2 \geqslant \| C^T dX^2 \|^2\end{array}\right\} \tag{2-4-15}$$

将此式代入(2-4-13)式,得固有非线性性和参数效应非线性性如下.

定义2-4-2 非线性模型(2-1-4)式的固有非线性性和参数效应非线性性分别定义为

$$\left.\begin{aligned}\alpha^N=\frac{\| C^N\mathrm{d}X^2 \|^2}{V'V}<1\\ \alpha^T=\frac{\| C^T\mathrm{d}X^2 \|^2}{V'V}<1\end{aligned}\right\} \tag{2-4-16}$$

对于任一非线性模型，当(2-4-16)式成立时，说明非线性模型的固有非线性性和参数效应非线性性都较弱，此非线性模型可以线性近似，且线性近似所引起的模型误差向量的平方和小于等于残差向量的平方和. 当(2-4-16)式不成立时，同样可能出现三种情况：第一种情况，$\alpha^N>1$，$\alpha^T>1$. 说明此非线性模型的固有非线性性和参数效应非线性性都很强，即该非线性模型的非线性强度很强，此非线性模型不能线性近似. 第二种情况，$\alpha^N>1$，$\alpha^T<1$. 说明此非线性模型的固有非线性性较强，只是参数选择得较好. 由于非线性模型的固有非线性性较强，所以此非线性模型仍不能线性近似. 第三种情况，$\alpha^N<1$，$\alpha^T>1$. 说明此非线性模型的固有非线性性较弱，只要设法对参数进行变换，使得在新参数下 $\alpha^T<1$，就可以在此新参数下对非线性模型进行线性近似.

例 2-4-3 根据(2-4-16)式判断例 2-1-1 和例 2-1-2 中的两个非线性模型能否线性近似. 例 2-4-2 已算出 $C\mathrm{d}X^2$ 和 $V'V$，再由例 2-2-1 知

$$S_T=\begin{pmatrix}0.8030 & 0.3401 & 0.0632 & -0.0931 & -0.1726\\ 0.3401 & 0.2610 & 0.2000 & 0.1531 & 0.1171\\ 0.0632 & 0.2000 & 0.2617 & 0.2779 & 0.2682\\ -0.0931 & 0.1531 & 0.2779 & 0.3278 & 0.3331\\ -0.1726 & 0.1171 & 0.2682 & 0.3331 & 0.3464\end{pmatrix}$$

于是

$$\begin{aligned}C^T\mathrm{d}X^2 &= S_TC\mathrm{d}X^2\\ &=(0.0017 \quad 0.0390 \quad 0.0569 \quad 0.0629 \quad 0.0620)'\\ C^N\mathrm{d}X^2 &= C\mathrm{d}X^2-C^T\mathrm{d}X^2\\ &=(0.0078 \quad -0.0101 \quad -0.0083 \quad 0.0013 \quad 0.0125)'\end{aligned}$$

故

$$\frac{\alpha^N \| C^N\mathrm{d}X^2 \|^2}{V'V}=\frac{0.0004}{0.000093}=4.1901>1$$

$$\frac{\alpha^T \| C^T\mathrm{d}X^2 \|^2}{V'V}=\frac{0.0126}{0.000093}=135.0743>1$$

此例的非线性模型固有非线性性较强，参数效应非线性性更强，故不能线性化.

再看例 2-1-2 中的非线性模型，由于 $V'V$ 和 $C\mathrm{d}X^2$ 已由例 2-4-2 算出，而由例 2-2-2 有

$$S_T=\begin{pmatrix}0.9042 & 0.2571 & 0.0415 & -0.0664 & -0.1311\\ 0.2571 & 0.2050 & 0.1876 & 0.1789 & 0.1737\\ 0.0415 & 0.1876 & 0.2363 & 0.2606 & 0.2753\\ -0.0664 & 0.1789 & 0.2606 & 0.3015 & 0.3260\\ -0.1311 & 0.1737 & 0.2753 & 0.3260 & 0.3556\end{pmatrix}$$

故：$C^T\mathrm{d}X^2=S_TC\mathrm{d}X^2=(0.0008 \quad 0.0004 \quad 0.0003 \quad 0.0002 \quad 0.0002)'$

$$C^N \mathrm{d}X^2 = C\mathrm{d}X^2 - C^T \mathrm{d}X^2 = (0 \quad 0 \quad 0 \quad 0 \quad 0)'$$

$$\alpha^N = \frac{\| C^N \mathrm{d}X^2 \|^2}{V'V} = 0 < 1$$

$$\alpha^T = \frac{\| C^N \mathrm{d}X^2 \|^2}{V'V} = \frac{9.381 \times 10^{-7}}{3.9 \times 10^{-5}} = 0.0241 < 1$$

所以该非线性模型在此观测精度下可以线性化.

用(2-4-16)式作为非线性模型能否线性近似的判据与(2-4-9)式相比,具有如下几个方面的优点:

1. 计算简单,由以上算例知,计算非线性模型的固有非线性性 α^N 和参数效应非线性性 α^T 比计算最大相对固有曲率 Γ^N 和最大相对参数效应曲率 Γ^T 简单得多.

2. 不需进行假设检验,而且判断准确,结论惟一. 因为 α^N 和 α^T 不是小于 1 就是大于等于 1,不可能有第三种取值. 而用(2-4-9)式判断,实际上是做 F 检验,而 F 检验不仅与分子自由度 t、分母自由度 $n-t$ 有关,还与显著水平 α 有关. 而 α 的选取有较大的随意性. 选不同的 α,可能会得到截然相反的结论. 另外,观测值的个数 n 对判断结果的影响较大,同一个非线性模型,当观测值的个数 n 不同时,结论也可能不同.

3. 当判断某非线性模型可以线性近似时,还可以估计线性近似后所引起的模型误差. 若按(2-4-9)式判断,则做不到这一点.

4. 当(2-4-16)式不成立时,同(2-4-9)式一样,同样可以分三种情况明确回答不成立的原因,以及能否通过参数变换再线性化. 即(2-4-16)式同样具备(2-4-9)式的特点.

第三章 非线性最小二乘估计

第一节 非线性最小二乘估计的定义及存在性定理

非线性模型(2-1-4)式相应的误差方程为:

$$V=f(\hat{X})-L \tag{3-1-1}$$

于是残差平方和为

$$V'V=\|V\|^2=\|f(\hat{X})-L\|^2=(f(\hat{X})-L)'(f(\hat{X})-L) \tag{3-1-2}$$

测量中,观测值常常有不同精度,甚至还有相关观测值.但根据等价观测理论[47](黄维彬,1992),这些观测值都可以变换为独立观测值.所以采用同精度观测讨论问题.

定义 3-1-1 非线性模型(2-1-4)式中参数 X 的一个估计量 $\hat{X}$,若满足下列关系:

$$V'V=\min \tag{3-1-3}$$

则称 $\hat{X}$ 是 X 的一个非线性最小二乘估计,用 $\hat{X}_{NLS}$ 表示.在不引起混淆的地方简记为 $\hat{X}$.其中 V 由(3-1-1)式确定.

这个定义与线性模型最小二乘估计的定义是完全一致的.(3-1-3)式的几何意义就是观测空间至解空间的距离最短,或者说 $f(\hat{X})$ 是解轨迹 π 上离观测值 L 最近的点(见图2-2-1).L 到 π 的距离就是 $\|V\|$.

根据定义 3-1-1,有如下定理:

定理 3-1-1 在非线性模型(2-1-4)式中,若 $f(X)$ 在参数空间 $\mathscr{X}$ 上关于 X 存在一阶连续偏导数,且 X 的非线性最小二乘估计量 $\hat{X}$ 存在,则残差向量 V 在 $\hat{X}$ 处垂直于切空间 T(见图 2-2-1).

证明: 因为 $V'V=\|f(\hat{X})-L\|^2$ 在 $\hat{X}$ 处达到极小值,因此有

$$\left.\frac{\partial V'V}{\partial X}\right|_{X=\hat{X}}=0$$

即

$$\left.\frac{\partial V'V}{\partial X}\right|_{X=\hat{X}}=2V'\frac{\partial V}{\partial X}=2V'\left.\frac{\partial f(X)}{\partial X}\right|_{X=\hat{X}}=2V'B(\hat{X})=0$$

因为 $V'B(\hat{X})=0$,所以向量 V 与矩阵 $B(\hat{X})$ 正交.而 $\hat{X}$ 处的切空间 T 是由 $B(\hat{X})$ 的列向量生成的空间,所以 V 在 $\hat{X}$ 处垂直于切空间 T.

(3-1-3)式定义了非线性最小二乘估计量 $\hat{X}$.那么在参数空间 $\mathscr{X}$ 中,是否存在这样的量呢? 这就是关于非线性最小二乘估计量的存在性问题.下面的定理圆满地回答了这个问题.

定理 3-1-2 假设 $\mathscr{X}$ 为 $\mathbf{R}^t$ 上的紧子集,$f(X)$ 关于 X 在 $\mathscr{X}$ 上连续,则必存在 $\mathbf{R}^n$ 上的可测函数 $\hat{X}=\hat{X}(L)$,使得

$$\|f(\hat{X}(L))-L\|^2=\min_{X\in\mathscr{X}}\|f(X)-L\|^2,\ L\in\mathbf{R}^n$$

此定理的证明可分两个方面，其一存在极小值．关于这一点，由于 $\| f(X) - L \|^2$ 在 $\mathscr{X} \times \mathbf{R}^n$ 上关于X 和L 都是连续函数．而$\mathscr{X}$为紧子集，所以对于任意 L，$\| f(X) - L \|^2$ 在$\mathscr{X}$上必存在极小值 $\hat{X}(L)$．其二就是证明可测性．关于这一方面的证明，超出本书的范围，有兴趣的读者可参阅文献〔24〕(韦博成，1989)中的定理 1.1．

第二节　非线性最小二乘估计的近似解法

既然在$\mathscr{X}$上存在最小二乘估计是 $\hat{X}$，能使 $\| f(X) - L \|^2 = \min$，那么，如何求解这一最小二乘估计量呢？本节和下面几节就来讨论这个问题．

一、线性近似

当非线性模型(2-1-4)式的非线性强度较弱，使得(2-4-16)式成立时，可以将非线性模型在 X_0 处线性化，并用线性模型的求解理论来解算非线性模型(2-1-4)式，这也就是我们大家所熟悉的传统方法．即将非线性模型(2-1-4)式在 X_0 处线性化，得误差方程：

$$V = B(X_0)\mathrm{d}X - (L - f(X_0))$$

根据最小二乘原理可解得：

$$\mathrm{d}X = (B'(X_0)B(X_0))^{-1}B'(X_0)(L - f(X_0))$$

于是参数 X 的最小二乘估计量为：

$$\hat{X} = X_0 + \mathrm{d}X$$

第二章已经证明，当(2-4-16)式成立时，如此线性近似所引起的模型误差对最小二乘估计量 $\hat{X}$ 的影响小于等于观测误差对参数估值$\hat{X}$ 的影响．

二、参数变换

当非线性模型(2-1-4)式的固有曲率较小，而参数效应曲率较大，即当 $\alpha^N < 1$，$\alpha^T > 1$ 时，可设法先对非线性模型进行参数变换，使得非线性模型在新参数下有 $\alpha^T > 1$，或者在新参数下使参数效应立体阵为0，然后再在新参数下对非线性模型进行线性化．

如何进行参数变换才能保证非线性模型在新参数下的曲率立体阵为 0 呢？这个问题可由下列定理回答．

定理 3-2-1　如欲通过参数变换 $Y = \Psi^*(X)$，使非线性模型在新参数 Y 下的参数效应立体阵 $H^* = 0$，则函数 $\Psi^*(X)$应满足下列微分方程：

$$\left.\begin{aligned} \frac{\partial^2 \varphi^*(X)}{\partial X^2} &= \left[\frac{\partial \varphi^*(X)}{\partial X}\right][T] \\ T &= [(B'B)^{-1}B'][W] \end{aligned}\right\} \tag{3-2-1}$$

式中：矩阵 B 和立阵W 的定义同前．

该定理的证明参见文献〔24〕(韦博成，1989)．

求解微分方程(3-2-1)式几乎是不可能的[24](韦博成，1989)．因此，定理 3-2-1 只是从理论上指出了使参数变换后，在新参数下的参数效应立体阵等于零的一般方法．由于微分方程(3-2-1)式无法求解，所以定理 3-2-1 只有理论上的意义，在实际中无法使用．也就是说，对于一般的固有曲率较小、参数效应曲率较大的非线性模型，虽然从理论上讲可以先通过参

数变换，然后在新参数下线性近似，但在实际中却难以实现. 因此，不再过多地讨论参数变换的理论.

第三节 非线性最小二乘估计的迭代解法

对于非线性强度很强，即 $\Gamma^N > \frac{1}{\sqrt{F}}$，$\Gamma^T > \frac{1}{\sqrt{F}}$（或 $\alpha^N > 1$，$\alpha^T > 1$）的非线性模型，由于线性近似将产生大于观测误差的模型误差，所以一般采用迭代的方法求解.

根据定义 3-1-1 知，求非线性模型(2-1-4)式的最小二乘估计量，就是求参数 X 的估值 $\hat{X}$，使

$$\begin{aligned} V'V(\hat{X}) &= (f(\hat{X}) - L)'(f(\hat{X}) - L) \\ &= f'(\hat{X})f(\hat{X}) - 2f'(\hat{X})L + L'L = \min \end{aligned} \tag{3-3-1}$$

由于 $L'L$ 是一常量，所以(3-3-1)式等价于目标函数为

$$R(\hat{X}) = f'(\hat{X})f(\hat{X}) - 2f'(\hat{X})L = \min \tag{3-3-2}$$

的非线性无约束最优化问题.

因为 $f(\hat{X})$是 $\hat{X}$ 的非线性函数，所以对(3-3-2)式求一阶偏导数，并令其为零，得不到 $\hat{X}$ 的显表达式，故求不出 $\hat{X}$ 的解析解. 因此，我们只能设法寻找某一近似解 X^*，使

$$R(X^*) \leqslant R(\hat{X}) \tag{3-3-3}$$

成立. 寻找使(3-3-3)式成立的近似解 X^*，一般只有采用迭代的方法. 为此，下面介绍几种常用的迭代方法.

1. 牛顿法

设 $R(\hat{X})$的极小值 X^* 的一个近似值为 $X^{(k)}$，在 $X^{(k)}$附近将 $R(X^*)$展为泰勒级数，取至二次项得

$$\begin{aligned} R(X^*) &= R(X^{(k)} + \mathrm{d}X^{(k)}) \\ &= R(X^{(k)}) + g^{(k)}\mathrm{d}X^{(k)} + \frac{1}{2}\mathrm{d}X'^{(k)}G_k\mathrm{d}X^{(k)} = \min \end{aligned} \tag{3-3-4}$$

式中：

$$\begin{aligned} g^{(k)} &= (g_1^{(k)} \quad g_2^{(k)} \quad \cdots \quad g_t^{(k)}) \\ &= \left.\begin{pmatrix} \frac{\partial R}{\partial x_1} & \frac{\partial R}{\partial x_2} & \cdots & \frac{\partial R}{\partial x_t} \end{pmatrix}\right|_{X=X^{(k)}} \end{aligned} \tag{3-3-5}$$

$$G_k = \begin{pmatrix} \frac{\partial^2 R}{\partial x_1^2} & \frac{\partial^2 R}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 R}{\partial x_1 \partial x_t} \\ \frac{\partial^2 R}{\partial x_2 \partial x_1} & \frac{\partial^2 R}{\partial x_2^2} & \cdots & \frac{\partial^2 R}{\partial x_2 \partial x_t} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 R}{\partial x_t \partial x_1} & \frac{\partial^2 R}{\partial x_t \partial x_2} & \cdots & \frac{\partial^2 R}{\partial x_t^2} \end{pmatrix}_{X=X^{(k)}} = G_k' \tag{3-3-6}$$

G_k 称为$X^{(k)}$处的 Hessian 矩阵.

$$\mathrm{d}X^{(k)} = X^* - X^{(k)} \tag{3-3-7}$$

$g^{(k)}$是 $R(X)$在 $X^{(k)}$处的梯度方向.

由于 $X^{(k)}$ 是 X^* 的一个已知的近似值,故(3-3-4)式只是 $dX^{(k)}$ 的函数,为了求得使(3-3-4)式成立的 $dX^{(k)}$,将(3-3-4)式对 $dX^{(k)}$ 求偏导,并令其为零,得

$$g^{(k)} + dX'^{(k)}G_k = 0$$

移项后两边转置,顾及(3-3-6)式,得

$$G_k dX^{(k)} = -g'^{(k)} \tag{3-3-8}$$

当 G_k 非奇异时,由(3-3-8)式可解得使(3-3-4)式成立的 $dX^{(k)}$,即

$$dX^{(k)} = -G_k^{-1}g'^{(k)} \tag{3-3-9}$$

当 $dX^{(k)}$ 充分小时,$dX^{(k)}$ 能使(3-3-4)式成立.但由于 X^* 未知,故 $dX^{(k)}$ 不能充分小,需不断迭代,直至 $dX^{(k)}$ 充分小,其迭代公式为

$$\begin{aligned} X^{(k+1)} &= X^{(k)} + dX^{(k)} \\ &= X^{(k)} - G_k^{-1}g'^{(k)} \end{aligned} \tag{3-3-10}$$

(3-3-10)式就是牛顿迭代的基本公式,终止迭代条件

$$R(X^{(k+1)}) = R(X^{(k)}) \tag{3-3-11}$$

或

$$g^{(k)} = 0 \tag{3-3-12}$$

由于 $R(X^{(k)})$ 是一个绝对值较大的数,而 $g^{(k)}$ 的各元素的绝对值都很小,因此,由于计算机有效数字的限制,以(3-3-11)式作为迭代收敛条件比(3-3-12)式作为迭代收敛条件收敛得要快一些.

牛顿法的迭代步骤为:

(1)选取初值 $X^{(0)}$,并令 $k=0$.

(2)按(3-3-5)式计算梯度方向 $g^{(k)}$,若 $g^{(k)}=0$ 则转至(7).

(3)计算矩阵 G_k.

(4)解线性方程组(3-3-8)式,得 $dX^{(k)}$.

(5)按(3-3-10)式计算新的近似值 $X^{(k+1)}$.

(6)计算目标函数值 $R(X^{(k+1)})$,若 $R(X^{(k+1)}) \neq R(X^{(k)})$ 则转至(2)继续迭代.

(7)终止迭代,输出 $X^{(k+1)}$ 和 $R(X^{(k+1)})$,结束.

例 3-3-1 设 $X^{(0)} = (5.4 \quad -0.3)'$,用牛顿法求解例 2-1-1 中非线性模型参数的非线性最小二乘估计.

由例 2-1-1 知,目标函数为

$$\begin{aligned} R(X) &= f'(X)f(X) - 2f'(X)L \\ &= x_1^2 \sum_{i=1}^{5} e^{2ix_2} - 2x_1 \sum_{i=1}^{5} L_i e^{ix_2} \end{aligned}$$

$$g = \begin{pmatrix} \dfrac{\partial R}{\partial x_1} \\ \dfrac{\partial R}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2(x_1 \sum\limits_{i=1}^{5} e^{2ix_2} - \sum\limits_{i=1}^{5} L_i e^{ix_2}) \\ 2x_1(x_1 \sum\limits_{i=1}^{5} i e^{2ix_2} - \sum\limits_{i=1}^{5} iL_i e^{ix_2}) \end{pmatrix}$$

$$G_k = \begin{pmatrix} \dfrac{\partial^2 R}{\partial x_1^2} & \dfrac{\partial^2 R}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 R}{\partial x_2 \partial x_1} & \dfrac{\partial^2 R}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 2\sum\limits_{i=1}^{5} e^{2ix_2} & 2(2x_1 \sum\limits_{i=1}^{5} L_i e^{ix_2} - \sum\limits_{i=1}^{5} iL_i e^{ix_2}) \\ 2(2x_1 \sum\limits_{i=1}^{5} i e^{2ix_2} - \sum\limits_{i=1}^{5} iL_i e^{ix_2}) & 2x_1(2x_1 \sum\limits_{i=1}^{5} i e^{2ix_2} - \sum\limits_{i=1}^{5} i^2 L_i e^{ix_2}) \end{pmatrix}$$

将 $X^{(0)}$ 代入计算 $g^{(0)}$ 和 G_0 后，按以上迭代程序迭代，结果列于表 3-3-1.

表 3-3-1

k		1	2	3	4	5	6
$g^{(k)}$	$g_1^{(k)}$	−1.205024908	0.3991833382	0.02889398018	0.0001691624122	3.9492×10^{9}	-2.4012×10^{9}
	$g_2^{(k)}$	−17.1530503	7.037242713	0.4948407424	0.002938012264	2.4039×10^{-7}	-1.5569×10^{7}
$X^{(k)}$	$x_1^{(k)}$	5.333013265	5.41719809	5.422708003	5.442744565	5.422744593	5.422744582
	$x_2^{(k)}$	−0.2539145225	−0.2542573375	−0.2556634078	−0.2556720853	−0.2556720877	−0.2556720866
$R(X^{(k)})$		−40.21054702	−40.58524686	−40.63522342	−40.63549278	−40.63549281	−40.63549281

迭代 6 次后，有 $R(X^{(6)})=R(X^{(5)})=-40.63549281$，所以停止迭代，得 X 的非线性最小二乘解为

$$X^*=\begin{pmatrix}5.422744582\\-0.255672087\end{pmatrix}$$

由例 2-1-1 知

$$\Delta X=X^*-X=\begin{pmatrix}+0.002608395\\-0.001310197\end{pmatrix}$$

$$\|\Delta X\|=0.0029$$

由例 2-1-1 知，本迭代解与其真值的距离比线性近似解与其真值的距离要小一个数量级.

当初值取 $X^{(0)}=(5.4\quad -0.5)'$ 时，迭代发散，这说明牛顿法对初值很敏感.

2. 信赖域法

我们随后将会证明，牛顿法具有很快的收敛速度，但它总是局部收敛的. 因为牛顿法的基本思想是用二次函数

$$Q(X^{(k)})=R(X^{(k)})+g^{(k)}\mathrm{d}X+\frac{1}{2}\mathrm{d}X'^{(k)}G_k\mathrm{d}X^{(k)}$$

去逼近 $R(X^*)$. 只有当 $\mathrm{d}X^{(k)}$ 充分小时，$Q(X^{(k)})$ 才能很好地逼近 $R(X^{(k)})$. 因此，人们近年来提出了一种保证全局收敛的方法，来代替牛顿法中的线性搜索.

既然只有当 $\mathrm{d}X^{(k)}$ 充分小时，$Q(X^{(k)})$ 才能逼近 $R(X^{(k)})$，那么可以对 $\mathrm{d}X$ 加以限制，然后在限制条件下来寻求 $R(X^*)$ 的极小值. 这个思想相当于求解下列约束最优化问题：

目标函数

$$Q(X^{(k+1)})=R(X^{(k)})+g^{(k)}\mathrm{d}X^{(k)}+\frac{1}{2}\mathrm{d}X'^{(k)}G_k\mathrm{d}X^{(k)}=\min \qquad (3\text{-}3\text{-}13)$$

约束条件

$$\|\mathrm{d}X^{(k)}\|\leqslant h_k$$

式中：h_k 为一正数，它随迭代而变化.

约束条件 $\| dX^{(k)} \| \leqslant h_k$ 限制了 $dX^{(k)}$，使 $\| dX^{(k)} \|$ 不大于 h_k，这样 $\| dX^{(k)} \|$ 总在一个给定的小区域中活动. 这个区域是可信赖的，所以称该方法为信赖域法.

常数 h_k 取决于 $Q(X^{(k+1)})$ 对 $R(X^{(k+1)})$ 的逼近程度. 这个逼近程度可用下式来描述

$$r_k = \frac{R(X^{(k+1)})}{Q(X^{(k+1)})} \tag{3-3-14}$$

r_k 越接近于 1，$Q(X^{(k+1)})$ 对 $R(X^{(k+1)})$ 的逼近程度越好，于是

$$h_{k+1} = \begin{cases} \dfrac{\| dX^{(k)} \|}{4}, & r_k < 0.25 \\ 2\| dX^{(k)} \|, & r_k > 0.75 \\ h_k, & 0.25 < r_k < 0.75 \end{cases} \tag{3-3-15}$$

这样，可总结出信赖域法的迭代程序：

(1)选取初值 $X^{(0)}$ 和 h_0，$k=0$.

(2)按(3-3-5)式和(3-3-6)式计算梯度方向 $g^{(k)}$ 和矩阵 G_k，若 $g^{(k)}=0$，则转(7).

(3)按(3-3-9)式计算 $dX^{(k)}$，并检查 $\| dX^{(k)} \|$ 是否满足约束条件，若不满足，则采取适当方法对 $\| dX^{(k)} \|$ 予以压缩. 然后在区域 h_k 内求使 $Q(X^{(k+1)}) = \min$ 的 $dX^{(k)}$.

(4)计算 X^* 的新的近似值 $X^{(k+1)} = X^{(k)} + dX^{(k)}$.

(5)按(3-3-14)式计算 r_k，并按(3-3-15)式确定 h_{k+1}.

(6)检查 $R(X^{(k+1)}) = R(X^{(k)})$ 是否成立，若不成立，则转(2)继续迭代.

(7)终止迭代，输出 $X^{(k+1)}$ 和 $R(X^{(k)})$，结束.

例 3-3-2 设 $X_0 = (5.4 \quad -0.3)'$，$h_0 = 0.08$，用信赖域法求解例 2-1-1 中非线性模型参数的非线性最小二乘估计.

$R(X)$、g 和 G_k 的表达式同上例，迭代计算的结果列于表 3-3-2.

表3 3 2

k		1	2	3	4	5	6
$g^{(k)}$	$g_1^{(k)}$	-1.1205024908	0.1247630234	0.00548626096	0.000123053314	7.605186×10 5	4.700803×10 5
	$g_2^{(k)}$	-17.1530503	2.746427624	0.09609929786	0.002154167737	1.331179×10 3	8.228461×10 4
$X^{(k)}$	$x_1^{(k)}$	5.342193488	5.421482607	5.422716379	5.422727153	5.422733806	5.422737919
	$x_2^{(k)}$	-0.2446971417	-2553841042	-0.2556656317	-0.2556680975	-0.255669209	-0.255670562
$R(X^{(k)})$		-40.21054702	-40.62555285	-40.63548246	-40.6254928	-40.63549279	-40.63549279

迭代六次，有 $R(X^{(5)}) = R(X^{(6)})$，所以停止迭代，得

$$X^* = \begin{pmatrix} 5.422737919 \\ -0.255670562 \end{pmatrix}, \Delta X = \begin{pmatrix} 0.002601732 \\ -0.001308672 \end{pmatrix}$$

$$\| \Delta X \| = 0.0029$$

当初值 $X^0=(5.0\quad -0.4)'$ 时，牛顿法发散. 而本法经过 18 次迭代后仍收敛到 $X^*=\begin{pmatrix}5.4227436\\-0.2556720336\end{pmatrix}$，但当 $X^0=(5.0\quad -0.5)'$ 时发散. 这说明信赖域法也与初值有关，仍然是局部收敛，并不像想像的那样全局收敛.

3. 拟牛顿法

牛顿法基于二次模型

$$R(X^{(k)}+\mathrm{d}X^{(k)})\approx R(X^{(k)})+g^{(k)}\mathrm{d}X^{(k)}+\frac{1}{2}\mathrm{d}X^{(k)\prime}G_k\mathrm{d}X^{(k)}$$

当 R 的形式很复杂时，求 R 的二阶偏导数阵 G_k 将非常困难. 为了避免求二阶偏导数，我们考虑用一个仅包含一阶偏导数信息的对称矩阵 Q_k 去逼近 G_k，然后再按牛顿法予以迭代. 可见拟牛顿法与牛顿法的差别就是用 Q_k 代替 G_k.

拟牛顿法的关键是寻找一个只包含一阶偏导数信息的 Q_k 矩阵. 关于 Q_k 矩阵的确定，详见文献[48]（邓乃扬，1982），故在此不赘述. 此处介绍一种确定 Q_k 阵的新方法——数值法.

由(3-3-6)式并顾及(3-3-5)式有

$$G_k=\begin{pmatrix}\frac{\partial^2 R}{\partial x_1^2} & \frac{\partial^2 R}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 R}{\partial x_1\partial x_t}\\ \frac{\partial^2 R}{\partial x_2\partial x_1} & \frac{\partial^2 R}{\partial x_2^2} & \cdots & \frac{\partial^2 R}{\partial x_2\partial x_t}\\ \vdots & \vdots & & \vdots\\ \frac{\partial^2 R}{\partial x_t\partial x_1} & \frac{\partial^2 R}{\partial x_t\partial x_2} & \cdots & \frac{\partial^2 R}{\partial x_t^2}\end{pmatrix}=\begin{pmatrix}\frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_t}\\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_t}\\ \vdots & \vdots & & \vdots\\ \frac{\partial g_t}{\partial x_1} & \frac{\partial g_t}{\partial x_2} & \cdots & \frac{\partial g_t}{\partial x_t}\end{pmatrix}$$

根据多元函数偏导数的定义

$$\frac{\partial f(x_1,x_2,\cdots,x_t)}{\partial x_i}=\lim_{\Delta x_i\to 0}\frac{f(x_1,\cdots,x_i+\Delta x_i,\cdots,x_t)-f(x_1,\cdots,x_t)}{\Delta x_i}$$

知：

$$G_k=\begin{pmatrix}\lim\limits_{\mathrm{d}x_1\to 0}\frac{g_1(x_1+\mathrm{d}x_1,x_2,\cdots,x_t)-g_1(x)}{\mathrm{d}x_1} & \cdots & \lim\limits_{\mathrm{d}x_t\to 0}\frac{g_1(x_1,\cdots,x_t+\mathrm{d}x_t)-g_1(x)}{\mathrm{d}x_t}\\ \lim\limits_{\mathrm{d}x_1\to 0}\frac{g_2(x_1+\mathrm{d}x_1,x_2,\cdots,x_t)-g_2(x)}{\mathrm{d}x_1} & \cdots & \lim\limits_{\mathrm{d}x_t\to 0}\frac{g_2(x_1,\cdots,x_t+\mathrm{d}x_t)-g_2(x)}{\mathrm{d}x_t}\\ \vdots & & \vdots\\ \lim\limits_{\mathrm{d}x_1\to 0}\frac{g_t(x_1+\mathrm{d}x_1,x_2,\cdots,x_t)-g_t(x)}{\mathrm{d}x_1} & \cdots & \lim\limits_{\mathrm{d}x_t\to 0}\frac{g_t(x_1,\cdots,x_t+\mathrm{d}x_t)-g_t(x)}{\mathrm{d}x_t}\end{pmatrix}$$

令

$$g^{(k)}=\begin{pmatrix}\bar{g}_1^{(k)}\\ \bar{g}_2^{(k)}\\ \vdots\\ \bar{g}_t^{(k)}\end{pmatrix}=\begin{pmatrix}\bar{g}_1^{(k)}(x_1^{(k+1)},x_2^{(k)},\cdots,x_t^{(k)})\\ \bar{g}_2^{(k)}(x_1^{(k)},x_2^{(k+1)},x_3^{(k)},\cdots,x_t^{(k)})\\ \vdots\\ \bar{g}_t^{(k)}(x_1^{(k)},x_2^{(k)},\cdots,x_t^{(k+1)})\end{pmatrix}\tag{3-3-16}$$

则去掉极限后，得 G_k 的近似矩阵：

$$G_k \approx Q_k = \begin{pmatrix} \dfrac{\bar{g}_1^{(k)} - g_1^{(k)}}{\mathrm{d}x_1^{(k)}} & \dfrac{1}{2}\left(\dfrac{\bar{g}_1^{(k)} - g_1^{(k)}}{\mathrm{d}x_2^{(k)}} + \dfrac{\bar{g}_2^{(k)} - g_2^{(k)}}{\mathrm{d}x_1^{(k)}}\right) & \cdots & \dfrac{1}{2}\left(\dfrac{\bar{g}_1^{(k)} - g_1^{(k)}}{\mathrm{d}x_t^{(k)}} + \dfrac{\bar{g}_t^{(k)} - g_t^{(k)}}{\mathrm{d}x_1^{(k)}}\right) \\ & \dfrac{\bar{g}_2^{(k)} - g_2^{(k)}}{\mathrm{d}x_2^{(k)}} & \cdots & \dfrac{1}{2}\left(\dfrac{\bar{g}_2^{(k)} - g_2^{(k)}}{\mathrm{d}x_t^{(k)}} + \dfrac{\bar{g}_t^{(k)} - g_t^{(k)}}{\mathrm{d}x_2^{(k)}}\right) \\ * & & \ddots & \vdots \\ & & & \dfrac{\bar{g}_t^{(k)} - g_t^{(k)}}{\mathrm{d}x_t^{(k)}} \end{pmatrix} \tag{3-3-17}$$

$*$ 的意思为对称矩阵，下同.

用(3-3-17)式定义的对称矩阵 Q_k 既能较准确地逼近G_k，又只包含 R 的一阶偏导数信息，不需要求二阶偏导数，有了 Q_k 后，一切迭代均按牛顿法进行. 由于拟牛顿法一开始就要按(3-3-17)式计算 Q_k 阵，所以计算前除了给定 X 的初值$X^{(0)}$外，还必须给定 $\mathrm{d}X$ 的初值 $\mathrm{d}X^{(0)}$. $\mathrm{d}X^{(0)}$可以这样确定：当给定 $X^{(0)}$后，将 $X^{(0)}$减去一个很接近 $X^{(0)}$的向量 $\bar{X}^{(0)}$，则差值就是 $\mathrm{d}X^{(0)}$，即

$$\mathrm{d}X^{(0)} = X^{(0)} - \bar{X}^{(0)} \tag{3-3-18}$$

开始计算时，用 $\bar{X}^{(0)}$和 $\mathrm{d}X^{(0)}$即可.

例 3-3-3 用拟牛顿法求解例2-1-1中非线性模型参数的非线性最小二乘估计. 初值为 $X^{(0)} = (5.4 \quad -0.3)'$.

由于给定 $X^{(0)} = (5.4 \quad -0.3)'$，则取 $\bar{X}^{(0)} = (5.39 \quad -0.28)$，由(3-3-18)式得 $\mathrm{d}X^{(0)} = (0.01 \quad 0.02)'$. 取定 $\bar{X}^{(0)}$和 $\mathrm{d}X^{(0)}$后，按拟牛顿法迭代的结果列于表 3-3-3. 由表 3-3-3 可以看出，迭代 6 次后，有 $R(X^{(5)}) = R(X^{(6)})$，停止迭代，得

$$X^* = \begin{pmatrix} 5.422744560 \\ -0.255672085 \end{pmatrix}$$

$$\| \Delta x \| = 0.0029$$

与以上迭代解相同，这表明用数值法确定的 Q_k 阵，能很好地逼近 G_k 阵.

4. 最速下降法

以上三种方法，都需要计算 $R(X)$的二阶偏导数矩阵，最速下降法则只假定 $R(X)$在解 $\hat{X}$ 附近具有二阶连续偏导数，且此二阶偏导数矩阵的行列式大于、等于 0. 并不需要具体计算这个矩阵.

最速下降法的基本思想是基于这样的事实：目标函数 $R(\hat{X})$在沿 $X^{(k)}$处的梯度方向 $g^{(k)}$上数值增加最快. 我们现在是要求目标函数 $R(\hat{X})$的最小值，因此，若在寻找 $R(\hat{X})$的最小值点 X^*的过程中，沿 $R(\hat{X})$在 $X^{(k)}$处的负梯度方向上寻找，势必使 $R(\hat{X})$的数值下降最快. 所以称按负梯度方向搜寻 X^*的方法为最速下降法. 其迭代公式为

$$X^{(k+1)} = X^{(k)} + \mathrm{d}X^{(k)} = X^{(k)} - \lambda^{(k)} g'^{(k)} \tag{3-3-19}$$

式中：λ 为实数，称为步长；梯度方向 $g^{(k)}$按(3-3-5)式计算.

从迭代公式(3-3-19)式知，最速下降法的关键是确定步长 λ. 下面就来讨论步长 λ 的确定方法.

表 3-3-3

k		1	2	3	4	5	6
$g^{(k)}$	$g_1^{(k)}$	−0.221245771	0.856623242	−0.03115185858	−0.00384076635	−0.00010347425	2.358884×10^{-7}
	$g_2^{(k)}$	−2.965959852	12.93742451	−0.4936312734	−0.0575278579	−0.0015617973	3.739449×10^{-6}
$X^{(k)}$	$x_1^{(k)}$	5.451669534	5.420543661	5.422705929	5.422744651	5.422744561	5.422744560
	$x_2^{(k)}$	−0.256609692	−0.255545476	−0.2556687702	−0.2556720931	−0.2556720852	−0.2556720849
$R(X^{(k)})$		−40.62536846	−40.48402317	−40.63523365	−40.63548925	−40.6354928	−40.6354928
Q_k	Q_{11}	2.5022628	2.753000625	2.765194851	2.763778092	2.76373428	2.76373428
	Q_{12}	26.20604801	33.01991882	31.12556489	31.13656682	31.1394768	31.1394768
	Q_{22}	398.1482354	533.4314266	472.3603199	469.4625468	469.7401401	469.7401401
G_k	G_{11}	2.311619915	3.003695448	2.780051726	2.763839956	2.763740298	2.763740270
	G_{12}	21.2196711	35.54791966	31.443537	31.13982534	31.13799006	31.13798972
	G_{22}	289.8392513	546.7992839	475.1482883	469.7863468	469.7540942	469.7540898

我们希望按(3-3-19)式得到的新的近似值 $X^{(k+1)}$ 是 $R(\hat{X})$ 在 $g^{(k)}$ 方向上的最小值，即

$$R(X^{(k+1)})=\min(R(X^{(k)}-\lambda^{(k)}g'^{(k)})) \tag{3-3-20}$$

为了推导公式方便，仍写为 $V'V=\min$ 的形式，于是(3-3-20)式等价于

$$V'V(X^{(k+1)})=\min(V'V(X^{(k)}-\lambda^{(k)}g'^{(k)}))$$

将 $V'V(X^{(k+1)})$ 在 $X^{(k)}$ 处展为泰勒级数，取至二次项，但略去 V 对 X 的二阶偏导数，得

$$\begin{aligned}V'V(X^{(k+1)})&=V'V(X^{(k)}-\lambda^{(k)}g'^{(k)})=v_1^2(X^{(k)}-\lambda^{(k)}g'^{(k)})+v_2^2(X^{(k)}-\lambda^{(k)}g'^{(k)})\\&\quad+\cdots+v_n^2(X^{(k)}-\lambda^{(k)}g'^{(k)})\\&\approx v_1^2(X^{(k)})-2\lambda^{(k)}v_1(X^{(k)})\sum_{j=1}^{t}g_j^{(k)}\frac{\partial f_1}{\partial x_j}+(\lambda^{(k)})^2\sum_{j=1}^{t}g_j^{(k)}\sum_{l=1}^{t}g_l^{(k)}\frac{\partial f_1}{\partial x_j}\frac{\partial f_1}{\partial x_l}\\&\quad+v_2^2(X^{(k)})-2\lambda^{(k)}v_2(X^{(k)})\sum_{j=1}^{t}g_j^{(k)}\frac{\partial f_2}{\partial x_j}+(\lambda^{(k)})^2\sum_{j=1}^{t}g_j^{(k)}\sum_{l=1}^{t}g_l^{(k)}\frac{\partial f_2}{\partial x_j}\frac{\partial f_2}{\partial x_l}\\&\quad+\cdots+v_n^2(X^{(k)})-2\lambda^{(k)}v_n(X^{(k)})\sum_{j=1}^{t}g_j^{(k)}\frac{\partial f_n}{\partial x_j}+(\lambda^{(k)})^2\sum_{j=1}^{t}g_j^{(k)}\\&\quad\sum_{l=1}^{t}g_l^{(k)}\frac{\partial f_n}{\partial x_j}\frac{\partial f_n}{\partial x_l}\end{aligned}$$

$$= V'V(X^{(k)}) - 2\lambda^{(k)} V'(X^{(k)})B(X^{(k)})g'^{(k)} + (\lambda^{(k)})^2 \sum_{i=1}^{n} g_i^{(k)} W_i^{(k)} g_i'^{(k)} \tag{3-3-21}$$

式中：

$$V(X^{(k)}) = (f_1(X^{(k)}) - L_1 \quad f_2(X^{(k)}) - L_2 \quad \cdots \quad f_n(X^{(k)}) - L_n)' \tag{3-3-22}$$

$$B(X^{(k)}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_t} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_t} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_t} \end{pmatrix}_{X = X^{(k)}} \tag{3-3-23}$$

$$W_i^{(k)} = \begin{pmatrix} \left(\frac{\partial f_i}{\partial x_1}\right)^2 & \frac{\partial f_i}{\partial x_1}\frac{\partial f_i}{\partial x_2} & \cdots & \frac{\partial f_i}{\partial x_1}\frac{\partial f_i}{\partial x_t} \\ \frac{\partial f_i}{\partial x_2}\frac{\partial f_i}{\partial x_1} & \left(\frac{\partial f_i}{\partial x_2}\right)^2 & \cdots & \frac{\partial f_i}{\partial x_2}\frac{\partial f_i}{\partial x_t} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_i}{\partial x_t}\frac{\partial f_i}{\partial x_1} & \frac{\partial f_i}{\partial x_t}\frac{\partial f_i}{\partial x_2} & \cdots & \left(\frac{\partial f_i}{\partial x_t}\right)^2 \end{pmatrix}_{X = X^{(k)}} \tag{3-3-24}$$

为了确定满足(3-3-20)式的 λ，将(3-3-21)式对 λ 求导数，并令其为0，得

$$\lambda^{(k)} = \frac{V'(X^{(k)})B(X^{(k)})g'^{(k)}}{\sum_{i=1}^{n} g_i^{(k)} W_i^{(k)} g_i'^{(k)}} \tag{3-3-25}$$

化简后，得

$$\lambda^{(k)} = \frac{(f(X^{(k)}) - L)'B(X^{(k)})B'(X^{(k)})(f(X^{(k)}) - L)}{2(f(X^{(k)}) - L)'B(X^{(k)})B'(X^{(k)})B(X^{(k)})B'(X^{(k)})(f(X^{(k)}) - L)} \tag{3-3-26}$$

最速下降法的迭代步骤如下：

(1)选取初值 $X^{(0)}$.

(2)按(3-3-5)式计算梯度方向 $g^{(k)}$，如果 $g^{(k)} = 0$ 则转(7).

(3)按(3-3-23)式和(3-3-24)式计算矩阵 $B(X^{(k)})$ 和 $W_i^{(k)}$.

(4)按(3-3-25)式计算 $\lambda^{(k)}$.

(5)按(3-3-19)式计算新的近似值 $X^{(k+1)}$，并计算 $R(X^{(k+1)})$.

(6)若 $R(X^{(k)}) \neq R(X^{(k+1)})$ 则转(2).

(7)终止迭代. 输出 $X^{(k+1)}$、$R(X^{(k+1)})$，结束.

例3-3-4 用最速下降法求解例2-1-1中非线性模型参数的非线性最小二乘估计. 初值为 $X^{(0)} = (5.4 \quad -0.3)'$.

在此初值下，按上述迭代程序，经452次迭代，得

$$X^* = \begin{pmatrix} 5.418092476 \\ -0.255371788 \end{pmatrix}, R(X^*) = -40.63548513$$

$$\| \Delta X \| = 0.0023$$

前已述及当 $X^{(0)}=(5.0\quad -0.5)'$ 时，上述三种方法均发散. 而最速下降法经1 480次迭代，仍收敛至 $X^*=(5.418337589\quad -0.2553721503)'$，$R(X^{(*)})=-40.63548593$. 甚至当 $X^{(0)}=(1.4\quad -1.1)'$ 时，经过 1 490 次迭代，也能收敛至 $X^*=(5.416988982\quad -0.2552803529)'$，$R(X^{(*)})=-40.63548108$. 一般说来，最速下降法对任意初值都能收敛，但收敛速度并不像该方法的名称那样是最速的，相反它却是最慢的. 这是因为最速下降法在接近最小值点时会产生拉锯现象.

5. 高斯-牛顿法

以上介绍的几种方法，都是求目标函数 $R(X^*)=\min$ 的非线性最优化算法. 与我们测量平差中已掌握的方法相去甚远. 因此，测量平差中的一些软件都不能直接应用. 而高斯-牛顿法则不同，几乎可以完全不改变原测量平差程序.

高斯-牛顿法的基本出发点就是在初值 $X^{(0)}$ 处对非线性模型进行线性近似，并按传统的平差方法求出一次近似值 $X^{(1)}$，然后反复迭代，直至前后两次 $V'V$ 值相等. 迭代步骤如下：

假设非线性模型(2-1-4)式存在一阶连续偏导数，且参数 X 之间相互独立，则在近似值 $X^{(0)}$ 处线性化，得误差方程：

$$V=B(X^{(0)})\mathrm{d}X-(L-f(X^{(0)}))$$

根据最小二乘原理，有

$$X^{(1)}=X^{(0)}+(B'(X^{(0)})B(X^{(0)}))^{-1}B'(X^{(0)})(L\quad f(X^{(0)}))$$

求得 $X^{(1)}$ 后，再以 $X^{(1)}$ 为近似值迭代，其迭代公式为

$$X^{(k+1)}=X^{(k)}+(B'(X^{(k)})B(X^{(k)}))^{-1}B'(X^{(k)})(L-f(X^{(k)}))\qquad(3\text{-}3\text{-}27)$$

终止迭代条件

$$R(X^{(k+1)})=R(X^{(k)}).$$

高斯-牛顿法具有一定的合理性. 因为若(2-1-4)式是线性模型，则有 $B(X^{(0)})=B$，$f(X^{(0)})=BX^{(0)}$. 于是

$$\begin{aligned}X^{(1)}&=X^{(0)}+(B'B)^{-1}B'(L-BX^{(0)})\\&=(B'B)^{-1}B'L\end{aligned}$$

上式表明：若(2-1-4)式是线性模型，则由高斯-牛顿法从任意初值出发，经一次迭代就可得到最小二乘估计的精确解. 当非线性模型(2-1-4)式的非线性强度较弱时，高斯-牛顿法是较好的方法.

例3-3-5 设 $X^{(0)}=(5.4\quad -0.3)'$，用高斯-牛顿法求解例 2-1-1 中非线性模型参数的非线性最小二乘估计.

按(3-3-26)式迭代的结果列于表 3-3-4.

表3-3-4

k		1	2	3	4	5
$X^{(k)}$	$x_1^{(k)}$	5.394141331	5.422298989	5.422744502	5.422744573	5.422744573
	$x_2^{(k)}$	−0.250050	−0.255618	−0.255672	−0.255672086	−0.255672086
$R(X^{(k)})$		−39.78568664	−40.62829761	−40.63549238	−40.6354928	−40.6354928

$$X^* = \begin{pmatrix} 5.422744573 \\ -0.255672086 \end{pmatrix},\ \| \Delta X \| = 0.0029$$

当 $X^{(0)} = (3.4 \quad -0.8)'$时，迭代发散．这说明虽然高斯-牛顿法有一定的合理性，但在具体执行时可能会产生一些问题．首先是对初值的依赖性较大．当初值较差时，会出现迭代发散现象，使迭代无法进行下去．好在我们在实际计算时，总是用观测值算出 $X^{(0)}$．$X^{(0)}$与 X 很接近，故一般可迭代收敛．

6．改进的高斯-牛顿法

高斯-牛顿法对初值的依赖性较强，当初值较差时易发散．为了克服这个缺点，下列定理给出了对高斯-牛顿法进行改进的思路[24]（韦博成，1989）．

定理 3-3-1 设 $X^{(k)}$是 X^* 的近似值，则 $V'V(X^{(k)})$一定不能达到最小，于是有

$$\left.\frac{V'V(X)}{\partial X}\right|_{X=X^{(k)}} \neq 0$$

设 $\mathrm{d}X^{(k)} = (B'(X^{(k)})B(X^{(k)}))^{-1}B'(X^{(k)})(L-f(X^{(k)}))$，那么必存在 $\lambda^* > 0$，使 $\lambda \in [0, \lambda^*]$时，有

$$V'V(X^{(k)} + \lambda \mathrm{d}X^{(k)}) < V'V(X^{(k)}) \tag{3-3-28}$$

证明：

将 $V'V(X^{(k)} + \lambda \mathrm{d}X^{(k)})$展为泰勒级数，得

$$V'V(X^{(k)} + \lambda \mathrm{d}X^{(k)}) = V'V(X^{(k)}) + \left.\frac{\partial V'V}{\partial X}\right|_{X=X^{(k)}} \lambda \mathrm{d}X^{(k)} + \alpha\lambda^2$$

因为

$$\begin{aligned} \frac{\partial V'V}{\partial X}\mathrm{d}X &= -2V'\frac{\partial f}{\partial X}\mathrm{d}X \\ &= -2(L-f(X))'B(X)(B'(X)B(X))^{-1}B'(X)(L-f(X)) \\ &= -2\| S_T(L-f(X)) \|^2 - V'V(X^{(k)}) \end{aligned}$$

所以

$$\begin{aligned} & V'V(X^{(k)} + \lambda \mathrm{d}X^{(k)}) - V'V(X^{(k)}) \\ &= (-2\| S_T(L-f(X)) \|^2 + \alpha\lambda)\lambda \end{aligned} \tag{3-3-29}$$

当 $\lambda \to 0$ 时，$\alpha \to 0$，而 $\alpha\lambda$ 是比 λ 更高阶的无穷小量，所以上式右端必为负值，因此(3-3-28)式得证．

此定理说明，当我们用高斯-牛顿法求出 $\mathrm{d}X^{(k)}$后，若适当选取 $\lambda^{(k)}$，使

$$X^{(k+1)} = X^{(k)} + \lambda^{(k)}\mathrm{d}X^{(k)} \tag{3-3-30}$$

则一定有：$V'V(X^{(k-1)}) < V'V(X^{(k)})$．这样就能保证 $V'V(X^{(k)})$逐步向 $V'V$ 的极小值靠近．于是就可避免迭代过程的波动性，从而保证得到收敛的非线性最小二乘估计．根据这个思想，可构成如下迭代算法：

(1) 根据初值 $X^{(0)}$，对非线性模型(2-1-4)式线性化，并按线性的最小二乘估计求 $\mathrm{d}X^{(0)}$．

(2) 适当选取 $\lambda^{(0)}$，并按(3-3-30)式计算 $X^{(1)}$．同时计算 $R(X^{(1)})$．

(3) 若 $R(X^{(1)}) = R(X^{(0)})$，则转(4)．否则令 $X^{(0)} = X^{(1)}$，转(1)．

(4) 输出 $X^{(1)}$和 $R^{(1)}$，结束.

该算法的关键是要计算 $\lambda^{(k)}$，韦博成建议对 $R(X)$采用三点抛物线近似[24]（韦博成，1989）. 即分别求出 $\lambda=0,\lambda=0.5$ 和 $\lambda=1$ 时 $R(X)$的值. 这三个值为：$R(X^{(k)})$，$R(X^{(k)}+\frac{1}{2}\mathrm{d}X^{(k)})$，$R(X^{(k)}+\mathrm{d}X^{(k)})$，则

$$\lambda^{(k)}=\frac{\frac{1}{2}+\frac{1}{4}(R(X^{(k)})-R(X^{(k)}+\mathrm{d}X^{(k)}))}{R(X^{(k)})-2R\left(X^{(k)}+\frac{1}{2}\mathrm{d}X^{(k)}+R(X^{(k)}+\mathrm{d}X^{(k)})\right)} \tag{3-3-31}$$

例 3-3-6 取初值 $X^{(0)}=(5.4 \quad -0.3)'$，用改进的高斯-牛顿法求解例 2-1-1 中非线性模型参数的非线性最小二乘估计.

迭代结果列于表 3-3-5.

表 3-3-5

k		1	2	3	4
$X^{(k)}$	$x_1^{(k)}$	5.394427485	5.42263568	5.422699198	5.42273328
	$x_2^{(k)}$	−0.252490	−0.255665	−0.255669	−0.255671
$R(X^{(k)})$		−39.78568664	−40.6341258	−40.63549279	−40.63549279
$\lambda^{(k)}$		0.9511570916	1.00123933	0.583333333	0.75

$$X^*=\begin{pmatrix}5.42273328\\-0.25567100\end{pmatrix},\ \|\Delta X\|=0.0029$$

当 $X^{(0)}=(3.4 \quad -0.8)'$时，高斯-牛顿法发散，而改进的高斯-牛顿法迭代 7 次收敛到 $X^*=(5.422744573 \quad -0.255672)'$，$R(X)=-40.6354928$，当 $X^{(0)}=(2.4 \quad -0.9)'$时，经 7 次迭代也能收敛到 $X^*=(5.422741415 \quad -0.255671)'$，$R(X)=-40.6354928$. 甚至当 $X^{(0)}=(8 \quad -1.5)'$时，经 10 次迭代也收敛到 $X^*=(5.422744573 \quad -0.255672086)'$，$R(X)=-40.6354928$. 这说明改进的高斯-牛顿法除具备高斯-牛顿法的全部优点外，还在很大程度上克服了高斯-牛顿法强烈依赖初值的缺点，是一种很有实用价值的方法.

7. 阻尼最小二乘法

高斯-牛顿法和改进的高斯-牛顿法有解的必要条件是 $B(X^{(k)})$矩阵列满秩. 但在非线性秩亏自由网平差中，由于缺少基准，$B(X^{(k)})$矩阵总是列降秩的. 在这种情况下，高斯-牛顿法和改进的高斯-牛顿法都不能用. 另外，当非线性模型(2-1-4)式中存在复共线关系时，尽管 $B(X^{(k)})$矩阵列满秩，但 $B'(X^{(k)})B(X^{(k)})$的条件数很大，使 $B'(X^{(k)})B(X^{(k)})$呈病态[49]（王德人，1979）. 这时高斯-牛顿法和改进的高斯-牛顿法也无法应用. 为了克服高斯-牛顿法和改进的高斯-牛顿法的这个致命弱点，可以采用增大 $B'(X^{(k)})B(X^{(k)})$的主对角线元素的办法，即将(3-3-27)式改写为

$$X^{(k+1)}=X^{(k)}+(B'(X^{(k)})B(X^{(k)})+\alpha^{(k)}I)^{-1}B'(X^{(k)})(L-f(X^{(k)})) \tag{3-3-32}$$

式中：$\alpha^{(k)}$为大于等于 0 的任意常数，称为阻尼因子. 因此称按(3-3-31)式迭代的算法为阻

尼最小二乘法. 该方法最早由 Levenberg 于 1944 年提出[50](Levenberg, 1944). 但当时并未受到人们的重视. 后来 Marquardt 又重新提出[51](Marquardt, 1963), 并进行了理论上的探讨. 所以人们又称阻尼最小二乘法为 LM 算法. 1971 年, Fletcher 又对其实现策略进行了改进[52](Fletcher, 1971), 改进后的算法又称 LMF 算法[53](熊西文等, 1986). 显然, 引进阻尼因子后, 矩阵$(B'(X^{(k)})B(X^{(k)})+\alpha^{(k)}I)$对任何正数 α, 总具有对称正定的性质. 因为此时矩阵$(B'(X^{(k)})B(X^{(k)})+\alpha^{(k)}I)$的所有特征值恒为正.

如果在(3-3-31)式的基础上再引进步长因子 $\lambda^{(k)}$, 则(3-3-31)式可写为:

$$X^{(k+1)}=X^{(k)}-\lambda^{(k)}(B'(X^{(k)})B(X^{(k)})+\alpha^{(k)}I)^{-1}V^{(k)} \tag{3-3-33}$$

式中: $V^{(k)}=f(X^{(k)})-L$; $\lambda^{(k)}$由(3-3-26)式确定. 关于 $\alpha^{(k)}$的确定[49](王德人和熊西文等)有较多的讨论. 我国著名的平差专家刘大杰教授针对测量平差给出了具体的选择策略和计算公式[33](刘大杰, 黄加纳, 1987). 对于非线性秩亏自由网平差问题, 根据我们的计算实践, 建议这样选取 $\alpha^{(k)}$:

$$\alpha^{(k)}=\mathrm{tr}(B'(X^{(k)})B(X^{(k)}))\sqrt{k} \tag{3-3-34}$$

然后根据 $V^{(k)}$和 $V^{(k+1)}$来调整 $\alpha^{(k+1)}$, 即

$$\alpha^{(k+1)}=\alpha^{(k)}\frac{V'V^{(k+1)}}{V'V^{(k)}} \tag{3-3-35}$$

适用于非线性秩亏自由网平差的阻尼最小二乘法的具体算法如下:

(1) 选取初值 $X^{(k)}$(此时 $k=0$), 计算 $V'V^{(k)}$.

(2) 将非线性模型(2-1-4)式在 $X^{(k)}$处线性化.

(3) 按一般间接平差组成法方程. 若 $k=0$, 则计算 $\alpha^{(k)}=\mathrm{tr}(B'^{(k)}B^{(k)})$.

(4) 按(3-3-26)式计算 $\lambda^{(k)}$.

(5) 按(3-3-32)式计算 $X^{(k+1)}$, 并计算 $V'V^{(k+1)}$. 若 $V'V^{(k+1)}=V'V^{(k)}$, 转(7).

(6) 按(3-3-34)式计算 $\alpha^{(k+1)}$, 转(2).

(7) 输出 $X^{(k+1)}$, $V'V^{(k+1)}$. 结束.

例 3-3-7 测边网如图3-3-1所示[54](高士纯、于正林, 1983), 同精度观测值和各点的近似坐标分别列于表 3-3-6.

表 3-3-6

序号	近似坐标		观测值
	X_i(m)	Y_i(m)	L_i(m)
1	0	0	22 141.335
2	22 141.335	0	27 908.063
3	19 187.335	20 265.887	20 044.592
4	−10 068.386	17 332.434	36 577.034
5			20 480.046
6			29 402.438

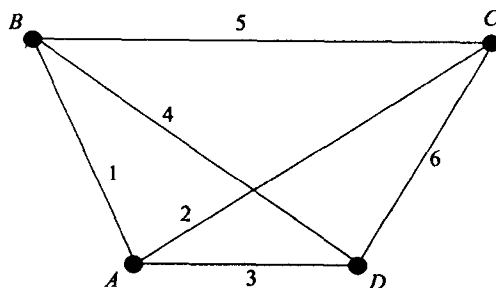


图 3-3-1

用阻尼最小二乘法求参数的最小二乘估计 $\hat{X}$.

按上述算法,迭代 9 次,收敛到

$$\hat{X}=(1.5427\times10^{-5}\quad 3.2685\times10^{-5}\quad 22141.33408\quad -1.1531\times10^{-6}\quad 19187.34327\quad 20265.88658\quad -10068.39257\quad 17332.43055)'$$

$$V'V^{(k+1)}=0.302914$$

而按秩亏自由网平差[54](高士纯、于正林),得

$$\hat{X}=(0.00117\quad 0.00379\quad 22141.33345\quad -0.00001\quad 19187.34260\quad 20265.88669\quad -10068.39322\quad 17332.43053)'$$

$$V'V=0.37087828$$

例 3-3-8 测角网如图3-3-2所示[54](高士纯、于正林, 1983),各点的近似坐标和同精度角度观测值列于表 3-3-7,用阻尼最小二乘法求参数的最小二乘估计 $\hat{X}$.

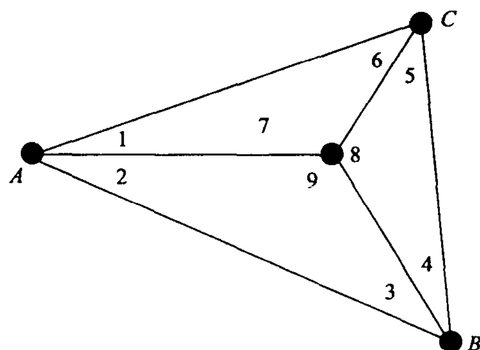


图 3-3-2

表 3-3-7

序号		1	2	3	4	5	6	7	8	9
近似坐标(m)	x	6 613.389	10 601.272	14 781.825	10 669.394					
	y	63 471.447	69 734.761	65 325.492	66 477.321					
观测值(m)		23°45′13.4″	20°58′22.7″	33°40′52.6″	42°16′38.9″	30°52′42.5″	28°26′05.6″	127°48′39.0″	106°50′41.2″	125°20′39.8″

用阻尼最小二乘法,迭代 13 次,得参数 X 的最小二乘估计 $\hat{X}$ 和 $V'V$ 为

$$\hat{X}=(6\,613.461803\quad 63\,471.44633\quad 10\,601.27786\quad 69\,734.80296\quad 14\,781.87943\quad 65\,325.47980\quad 10\,669.38083\quad 66\,477.32210)'$$

$$V'V^{(13)}=22.59345631$$

用秩亏自由网平差[54](高士纯、于正林),得: $V'V=22.63057$

$$\hat{X}=(6\,613.431\quad 63\,471.457\quad 10\,601.259\quad 69\,734.792\quad 14\,781.839\quad 65\,325.458\quad 10\,669.351\quad 66\,477.314)'$$

由此两例知,线性近似达不到 $V'V$ 的最小值.

第四节　几种迭代算法的效率比较

一、迭代过程的收敛性与收敛速率的有关定义

上节介绍了几种迭代算法. 它们的共同之处是通过各种不同的迭代算法产生迭代序列 $\{X^{(k)}\}\subset\mathbf{R}^n$. 不同的迭代算法, 有的强烈依赖于初值, 有的与初值关系不很密切; 有的收敛很快, 有的收敛却很慢. 之所以会这样, 是因为不同的算法具有不同的收敛性和不同的收敛速率. 关于收敛性和收敛速率, 有如下定义:

定义 3-4-1　若一算法对于某类目标函数来说, 任意给出初始点 $X^{(0)}\in\mathbf{R}^t$, 按该算法构造的序列 $\{X^{(k)}\}$ 总停止或者收敛到目标函数的一个极小点, 则称该算法对该类函数具有全局收敛性.

定义 3-4-2　若一算法对于某类目标函数来说, 在其定义域的某个区域 D 上任取一点作为初始点 $X^{(0)}$, 按该算法构造的序列 $\{X^{(k)}\}$ 属于区域 D, 并总停止或收敛到这个区域内的一个极小点, 则称该算法对该类目标函数具有区域 D 上的局部收敛性.

定义 3-4-3　若对于序列 $\{X^{(k)}\}$, 存在着 $P\geqslant 0$, 并存在常数 N 和 M, 使得当 $R\geqslant N$ 时, 有

$$\| X^{(k+1)}-X^{*}\| \leqslant M\| X^{(k+1)}-X^{*}\|^{P}$$

则称序列 $\{X^{(k)}\}$ 收敛的级数是 P, 或者说 $\{X^{(k)}\}$ 为 P 级收敛.

定义 3-4-4　若对于序列 $\{X^{(k)}\}$ 来说, 存在着常数 N 和 M, 使得当 $k\geqslant N$ 时, 有

$$\| X^{(k+m)}-X^{*}\| \leqslant M\| X^{(k)}-X^{*}\|^{2}$$

则称序列 $\{X^{(k)}\}$ 为 m 步二级收敛, 或 m 步平方收敛.

显然, 当 $m=1$ 时, 就是通常的平方收敛.

定义 3-4-5　若对于序列 $\{X^{(k)}\}$ 来说, 存在着 $\theta\in(0,1)$, 并存在常数 N 和 M, 使得当 $k\geqslant N$ 时, 有

$$\| X^{(k)}-X^{*}\| < M\theta^{k}$$

则称序列 $\{X^{(k)}\}$ 线性收敛.

一个实用的迭代算法至少应具有线性收敛速率.

定义 3-4-6　若对于序列 $\{X^{(k)}\}$ 来说, 任给 $\beta>0$, 都存在 $N>0$, 使当 $k\geqslant N$ 时, 有

$$\| X^{(k+1)}-X^{*}\| \leqslant\beta\| X^{(k)}-X^{*}\|$$

则称序列 $\{X^{(k)}\}$ 超线性收敛.

显然, 线性收敛是超线性收敛的必要条件, 而超线性收敛又是平方收敛的必要条件.

二、上述几种算法的收敛性与收敛速率

1. 牛顿法的收敛性与收敛速率

关于牛顿法的收敛性与收敛速率, 有如下定理:

定理 3-4-1　设目标函数 $R(X)=f'(x)f(x)-2f'(x)L$ 是某一开域内的三次连续可微函数, 且它在该开域内有极小点 X^{*}. 若存在着 $\varepsilon>0$ 和 $m>0$, 使得当

$$\| X-X^{*}\| \leqslant\varepsilon,\ Y\in\mathbf{R}^t$$

时,有

$$m\|Y\|^2\leqslant(Y,G(X)Y)$$

则当初始点 $X^{(0)}$ 充分接近 X^* 时,由牛顿法(3-3-10)式产生的序列$\{X^{(k)}\}$收敛于 X^*,且为平方收敛.

上式中 $G(X)$ 为 $R(X)$ 在点 X 处的 Hessian 矩阵,$(Y,G(X)Y)$ 表示两向量 Y 与 $G(X)Y$ 的内积.

证明: 显然定理条件中的不等式可加强为当 $\|X-X^*\|<\varepsilon$, $Y\in\mathbf{R}^t$ 时有

$$m\|Y\|^2\leqslant(Y,G(X)Y)\leqslant M\|Y\|^2$$

还可以看出,当 $\|X-X^*\|\leqslant\varepsilon$ 时,$G(X)$ 的最小特征值 $\lambda_1\geqslant m$.因而 $G^{-1}(m)$ 的最大特征值$\frac{1}{\lambda_1}=\|G^{-1}(X)\|\leqslant\frac{1}{m}$,因此

$$\|G^{-1}(X)g(X)\|\leqslant\|G^{-1}(X)\|\,\|g(X)\|\leqslant\frac{1}{m}\|g(X)\|$$

由 $g(X)$ 的连续性及 $g(X^*)=0$ 知,必存在着 $\varepsilon'\in\left(0,\frac{\varepsilon}{2}\right)$,使当 $\|X-X^*\|\leqslant\varepsilon'$ 时有

$$\|g(X)\|\leqslant\frac{m\varepsilon}{2}$$

这意味着当 $X^{(k)}$ 满足

$$\|X^{(k)}-X^*\|<\varepsilon'$$

时,$X^{(k+1)}$ 满足

$$\begin{aligned}\|X^{(k+1)}-X^*\|&=\|X^{(k)}-G^{-1}(X^{(k)})g(X^{(k)})-X^*\|\\&\leqslant\|X^{(k)}-X^*\|+\|G^{-1}(X^{(k)})g(X^{(k)})\|\\&\leqslant\frac{\varepsilon}{2}+\frac{1}{m}\cdot\frac{m\varepsilon}{2}=\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon\end{aligned}$$

这就证明了当 $X^{(0)}$ 充分接近 X^* 时,牛顿法产生的序列$\{X^{(k)}\}$收敛.根据定义 3-4-2 知,牛顿法对目标函数 $R(X)$ 局部收敛.

下面再来证明收敛速率.为此需要建立 $\|X^{(k)}-X^*\|$ 和 $\|X^{(k+1)}-X^*\|$ 之间的估计式.根据邓乃扬的推导[48],得

$$\|X^{(k+1)}-X^*\|\leqslant C\|X^{(k+1)}-X^*\|^2$$

所以根据定义 3-4-4,牛顿法的收敛速率是平方收敛.

信赖域法和拟牛顿法以及高斯-牛顿法,其实质是牛顿法,所以它们与牛顿法具有相同的收敛性质和收敛速率.

信赖域法和拟牛顿法的实质是牛顿法,这是显而易见的.那么,何以见得高斯-牛顿法的实质也是牛顿法呢?下面扼要说明这一点[55](陈宝林,1989).我们知道高斯-牛顿法是用线性函数 $\overline{V}=BX-L$ 去逼近非线性函数 $V=f(X)-L$.这相当于用目标函数

$$\Phi(X)=X'B'BX-2X'B'L \tag{3-4-1}$$

去逼近目标函数

$$R(X)=f'(X)f(X)-2f'(X)L$$

于是,根据牛顿法,我们很容易写出(3-4-1)式的牛顿迭代公式.由(3-4-1)式知,梯度方向为

$$g'^{(k)} = 2B'^{(k)}(B^{(k)}X^{(k)} - L) = 2B'^{(k)}(f(X^{(k)}) - L) \tag{3-4-2}$$

(顾及了用 $BX - L$ 去逼近 $f(X) - L$ 这一事实)，而 $\Phi(X)$ 在 $X^{(k)}$ 处的 Hessian 矩阵为

$$G_k = \frac{\partial^2 \varphi}{\partial X^2} = 2B'^{(k)}B^{(k)} \tag{3-4-3}$$

于是由(3-3-10)式得牛顿迭代公式

$$X^{(k+1)} = X^{(k)} - G_k^{-1}g'^{(k)} = X^{(k)} - (2B'^{(k)}B^{(k)})^{-1}(2B'^{(k)}(f(X^{(k)}) - L))$$
$$= X^{(k)} - (B'^{(k)}B^{(k)})^{-1}B'^{(k)}(f(X^{(k)}) - L) \tag{3-4-4}$$

(3-4-4)式正是高斯-牛顿法的迭代公式(3-3-27)式，所以说高斯-牛顿法的实质是牛顿法.

2. 最速下降法的收敛性与收敛速率

关于最速下降法的收敛性质与收敛速率，有如下定理：

定理 3-4-2 若目标函数 $R(X)$ 满足：

(i) $R(X)$ 是 $\mathbf{R}^t$ 上的二次连续可微函数；

(ii) 对于任意的 $X' \in \mathbf{R}^t$，存在着常数 $m > 0$，使当 $X \in C(X') = \{X \mid R(X)R(X')\}$，$Y \in \mathbf{R}^t$ 时，恒有

$$m\|Y\|^2 \leqslant (Y, G(X)Y)$$

(其中 $C(X')$ 为相对于 X' 的基准集，$G(X)$ 是 $R(X)$ 在 X 处的 Hessian 矩阵)，则从 $\mathbf{R}^t$ 中任意一点 $X^{(k)}$ 出发，当按最速下降法迭代时，或者在有限步内达到 $R(X)$ 的惟一极小点，或者所构造的序列 $\{X^{(k)}\}$ 收敛于 $R(X)$ 的惟一极小点.

定理 3-4-3 若目标函数 $R(X)$ 在 X^* 的某一邻域内二次连续可微，且存在 $\varepsilon > 0$ 和 $m > 0$，使当 $\|X - X^*\| < \varepsilon$，$Y \in \mathbf{R}^n$ 时，有

$$m\|Y\|^2 \leqslant (Y, G(X)Y)$$

则最速下降法构造的序列 $\{X^{(k)}\}$ 至少线性收敛于 X^*.

以上定理的证明，详见[48](邓乃扬，1982). 这两定理指出最速下降法是全局收敛的，且收敛速率是线性的.

3. 改进的高斯-牛顿法的收敛性及收敛速率

定理 3-4-4 设 $\mathscr{X}$ 是 $\mathbf{R}^t$ 上的有界凸集，X_0 是 $\mathscr{X}$ 的内点. 并且：

(i) $B(X) = \dfrac{\partial f}{\partial X}$ 在 $\bar{\mathscr{X}}$ 上存在且连续，而且是列满秩矩时，$\bar{\mathscr{X}}$ 是 $\mathscr{X}$ 的闭包.

(ii) $R(X_0) < \tilde{R}$ ($\tilde{R} = \min R(X_b)$，X_b 为 $\mathscr{X}$ 的边界点).

(iii) 在 $\mathscr{X}$ 上不存在二点 X' 和 X'' 使得下式成立：

$$R(X') = R(X''), \left.\frac{\partial R(X)}{\partial X}\right|_{X=X'} = \left.\frac{\partial R(X)}{\partial X}\right|_{X=X''}$$

则在这些条件下按改进的高斯-牛顿法得到的序列 $\{X^{(k)}\}$ 满足：

(i) 一切 $X^{(k)}$ 都是 $\mathscr{X}$ 的内点，$k = 1, 2, \cdots$

(ii) $\{X^{(k)}\}$ 收敛到极限 X^*，X^* 也是 $\mathscr{X}$ 的内点，而且是 $R(X)$ 在 $\mathscr{X}$ 上的稳定点，即有

$$\left.\frac{\partial R(X)}{\partial X}\right|_{X=X^*} = 0$$

证明：由于对于一切 $X^{(k)}$ 有 $R(X^{(k)}) \leqslant R(X_0) < \tilde{R}$，因此 $X^{(k)}$ 不可能是边界点($k = 1, 2, \cdots$)，所以结论(i)成立. 因为 $R(X^{(k)})$ 是 k 的单调递减序列，所以必有极限存在，设为

R^*. 又由于序列$\{X^{(k)}\}$有界,因此必存在收敛子序列,今设$\{X_{k_\beta}\}$, $\beta=1,2,\cdots$收敛于 X^*. 由函数$R(X)$的连续性可知:

$$\lim_{\beta\to\infty} R(X_{k_\beta}) = R(X^*) = R^* \leqslant R(X_0) < \tilde{R}$$

因此, X^* 亦为 $\mathscr{X}$ 的内点.

下面再来证明$\left.\dfrac{\partial R(X)}{\partial X}\right|_{X=X^*}=0$.

由定理条件(i)并顾及

$$\left(\left.\frac{\partial R(X)}{\partial X}\right|_{X=X^*}\right)' = -2B'(X^*)(L-f(X^*)) = -2B'(X^*)B(X^*)\mathrm{d}X^* \tag{3-4-5}$$

式中: $\mathrm{d}X^*=(B'(X^*)B(X^*))^{-1}B'(X^*)(L-f(X^*))$

知$\left.\dfrac{\partial R(X)}{\partial X}\right|_{X=X^*}=0$ 等价于 $\mathrm{d}X^*=0$. 为了证明 $\mathrm{d}X^*=0$, 用反证法, 即若 $\mathrm{d}X^*\neq 0$, 则推出矛盾结果.

由于 X^* 与 $\mathscr{X}$ 的内点, 故必存在充分小的 $0<\delta\leqslant 1$, 使 $\lambda\in[-\delta,\delta]$时 $X^*+\lambda\mathrm{d}X^*$ 为 $\mathscr{X}$ 的内点.

另由(3-3-29)式知, 存在 $0<\lambda^*\leqslant\frac{1}{2}\delta$, 使得

$$R(X^*+\lambda^*\mathrm{d}X^*)-R(X^*)<-C \tag{3-4-6}$$

其中 $C>0$, 但与 λ^* 有关. 又由于 $\beta\to\infty$ 时有:

$$X_\beta\to X^*,\ \mathrm{d}X_\beta\to\mathrm{d}X^*$$

取 $\beta(\lambda^*)$充分大, 使一切 $\beta>\beta(\lambda^*)$时 $X_\beta+\lambda^*\mathrm{d}X_\beta$ 亦为 $\mathscr{X}$ 的内点, 且有

$$R(X_\beta+\lambda^*\mathrm{d}X_\beta)-R(X^*+\lambda^*\mathrm{d}X^*)<\frac{C}{2} \tag{3-4-7}$$

(3-4-6)式成立是因为 $R(X)$在 $X^*+\lambda^*\mathrm{d}X^*$ 处连续. 因此顾及(3-4-6)式有

$$R(X_\beta+\lambda^*\mathrm{d}X_\beta)-R(X^*)<-\frac{C}{2}$$

由于一切 $X_\beta+\lambda^*\mathrm{d}X_\beta$ 都是 $\mathscr{X}$ 的内点, 因此

$$\lambda^*\in\Lambda_\beta=\{\lambda^*:0\leqslant\lambda^*\leqslant 1,\ X_\beta+\lambda^*\mathrm{d}X_\beta\in\tilde{\mathscr{X}}\}$$

故有

$$R(X_{k_{\beta+1}})=\min_{\Lambda_\beta} R(X_\beta+\lambda\mathrm{d}X_\beta)\leqslant R(X_\beta+\lambda^*\mathrm{d}X_\beta)$$

从而得到

$$R(X^*)\leqslant R(X_{k_{\beta+1}})\leqslant R(X_\beta+\lambda^*\mathrm{d}X_\beta)<R(X^*)-\frac{C}{2}$$

以上不等式是不可能成立的, 这样就证明了 $\mathrm{d}X^*=0$. 因此$\left.\dfrac{\partial R(X)}{\partial X}\right|_{X=X^*}=0$.

最后再来证明序列$\{X^{(k)}\}$的收敛性. 为此设$\{X^{(k)}\}$的另一收敛子序列为$\{\bar{X}_\beta\}$. 则根据以上同样的讨论知, 存在 $\bar{X}^*$, 当 $\beta\to\infty$, $\bar{X}_\beta\to\bar{X}^*$, $R(\bar{X}_\beta)\to R(\bar{X}^*)$. 且$\left.\dfrac{\partial R(X)}{\partial X}\right|_{X=\bar{X}^*}=0$. 由假设条件(iii)知, 必有 $\bar{X}^*=X^*$. 因此$\{X^{(k)}\}$必收敛, 且收敛到内点 X^*. 故定理得证.

至于收敛速率，也与高斯-牛顿法相同.

4.阻尼最小二乘法的收敛性与收敛速率

关于阻尼最小二乘法的收敛性，有下列定理.

定理3-4-5 设非线性模型(2-1-4)式中的 $f(X)$ 为定义在 $D\subset \mathbf{R}^n$ 上而取值于 $\mathbf{R}^t$ 的向量函数，它在包含 X^* 的区域 $D_0\subset D$ 上可微，并且对于任何 $X, Y\in D$，满足

$$\| B(X)-B(Y)\| \leqslant K\| X-Y\|$$

式中：K 为常数，且 $K>0$. 现令 $\lambda(X^*)$ 为矩阵 $B'(X^*)B(X^*)$ 的最小特征值，且

$$\lambda(X^*)>K\| f(X^*)\|$$

$\{\alpha_k\}$ 为任何非负有界实数序列，则由阻尼最小二乘法产生的序列 $\{X^{(k)}\}$ 局部收敛于 X^*. 证明详见文献[49](王德人，1979).

至于阻尼最小二乘法的收敛速率，有如下结论：随阻尼因子 α 的取值从 0 变化到无穷，其收敛速率由平方收敛减至线性收敛.

关于这一结论，我们可作如下简要说明，因为当 $\alpha=0$ 时，阻尼最小二乘法(3-3-32)式还原为高斯-牛顿法，即

$$X^{(k+1)}=X^{(k)}+(B'^{(k)}B^{(k)})^{-1}B'^{(k)}(L-f(X^{(k)}))$$

我们在前面已经证明，高斯-牛顿法具有平方收敛速率. 当 $\alpha\to\infty$ 时，(3-3-32)式的下降方向 $-B'^{(k)}(f(X^{(k)})-L)$ 趋近于最速下降法(3-3-19)式的负梯度方向 $-g'^{(k)}$(这点的证明详见文献[49](王德人，1979)). 故当 $\alpha\to\infty$ 时，阻尼最小二乘法趋近于最速下降法. 而定理3-4-3告诉我们，最速下降法为线性收敛. 所以阻尼最小二乘法随 α 从 0 变化到∞，收敛速率从平方收敛减至线性收敛. 这一点我们在实际计算中也得到了验证.

三、上述几种算法的效率比较

一般来讲，一个最好的迭代算法应该是一个最经济的方法，即能用最少时间求出满足精度要求的解 X^*[55](陈宝林，1989). 概括地讲，一种算法的效率如何，应从下列三个方面去综合衡量：

(1)迭代过程中每步的计算工作量.

(2)迭代序列的收敛性.

(3)迭代序列的收敛速率.

对于非线性模型的测量平差而言，还应考虑能否充分利用多年来所开发的大量平差软件. 因此，一个好的迭代算法，应该具有很快的收敛速率，应该对初值的依赖性不太强，而且每步的计算工作量都较小. 更重要的是还应该在任何情况下都能顺利完成迭代计算同时又能充分利用现有软件资源.

牛顿类算法(牛顿法、信赖域法、拟牛顿法等)尽管收敛快，但需要计算目标函数 $V'V$ 的 Hessian 矩阵，而且还要求 Hessian 矩阵正定. 不仅每一步的工作量都很大，而且都是局部收敛，加之与我们的平差习惯相去甚远，不便利用现有平差软件. 因此，在非线性模型平差中不能算好的算法.

最速下降法虽然具有全局收敛性，但收敛太慢，由例 3-3-4 知，同样的初值，收敛到同样的精度，牛顿法只要迭代 6 次，而最速下降法却需迭代 452 次.

高斯-牛顿法和改进的高斯-牛顿法，既具有牛顿法的收敛速率，每次迭代计算工作量又

不太大，尤其是改进的高斯-牛顿法，尽管也是局部收敛，但收敛域的范围相对较大. 还可充分利用现有平差软件，所以当非线性模型平差系统中具有足够基准时，使用改进的高斯-牛顿法有较高的效率. 当非线性模型平差系统中缺乏基准时，阻尼最小二乘法仍能顺利迭代，而阻尼最小二乘法亦不需太多地改变传统平差程序，所以，对于非线性模型秩亏自由网平差，采用阻尼最小二乘法为好.

第五节　非线性最小二乘估计的直接解法

前面介绍了几种迭代算法，下面再来介绍笔者提出的一种无需迭代的求解非线性模型参数的非线性最小二乘估计的直接解法[40]（王新洲，1999）.

一、基本原理

设非线性模型(2-1-4)式的三阶连续导数存在，并设参数 X 的近似值为 $X^{(0)}$，$X^{(0)}$的改正数为 dX，将(2-1-4)式展为泰勒级数，并取至三次项，得

$$L = f(X^{(0)}) + B\mathrm{d}X + \frac{1}{2}CY + \frac{1}{6}DZ + \Delta \tag{3-5-1}$$

相应的误差方程为

$$V = B\mathrm{d}X + \frac{1}{2}CY + \frac{1}{6}DZ + l \tag{3-5-2}$$

式中：$l - f(X^{(0)}) - L$；B，C 分别为 $n \times t$ 和 $n \times \alpha$ 的矩阵，其定义如前；D 为 $n \times \beta$ 的矩阵，由下式定义：

$$D = \begin{bmatrix} \frac{\partial^3 f_1}{\partial x_1^3} \cdots \frac{\partial^3 f_1}{\partial x_t^3} & 3\frac{\partial^3 f_1}{\partial x_1^2 \partial x_2} \cdots 3\frac{\partial^3 f_1}{\partial x_1^2 \partial x_t} & 3\frac{\partial^3 f_1}{\partial x_2^2 \partial x_1} \cdots 3\frac{\partial^3 f_1}{\partial x_t^2 \partial x_{t-1}} & 6\frac{\partial^3 f_1}{\partial x_1 \partial x_2 \partial x_3} \cdots 6\frac{\partial^3 f_1}{\partial x_{t-2} \partial x_{t-1} \partial x_t} \\ \frac{\partial^3 f_2}{\partial x_1^3} \cdots \frac{\partial^3 f_2}{\partial x_t^3} & 3\frac{\partial^3 f_2}{\partial x_1^2 \partial x_2} \cdots 3\frac{\partial^3 f_2}{\partial x_1^2 \partial x_t} & 3\frac{\partial^3 f_2}{\partial x_2^2 \partial x_1} \cdots 3\frac{\partial^3 f_2}{\partial x_t^2 \partial x_{t-1}} & 6\frac{\partial^3 f_2}{\partial x_1 \partial x_2 \partial x_3} \cdots 6\frac{\partial^3 f_2}{\partial x_{t-2} \partial x_{t-1} \partial x_t} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^3 f_n}{\partial x_1^3} \cdots \frac{\partial^3 f_n}{\partial x_t^3} & 3\frac{\partial^3 f_n}{\partial x_1^2 \partial x_2} \cdots 3\frac{\partial^3 f_n}{\partial x_1^2 \partial x_t} & 3\frac{\partial^3 f_n}{\partial x_2^2 \partial x_1} \cdots 3\frac{\partial^3 f_n}{\partial x_t^2 \partial x_{t-1}} & 6\frac{\partial^3 f_n}{\partial x_1 \partial x_2 \partial x_3} \cdots 6\frac{\partial^3 f_n}{\partial x_{t-2} \partial x_{t-1} \partial x_t} \end{bmatrix}_{X = X_0} \tag{3-5-3}$$

$$\alpha = \frac{1}{2}t(t+1), \beta = t^2 + C_t^3 \tag{3-5-4}$$

式中：C_t^3 表示 t 个量中拿出 3 个的组合

$$\mathrm{d}X = (\mathrm{d}x_1 \quad \mathrm{d}x_2 \quad \cdots \quad \mathrm{d}x_t)'$$

$$Y = \mathrm{d}X^2 = (\mathrm{d}x_1^2, \cdots, \mathrm{d}x_t^2, \mathrm{d}x_1\mathrm{d}x_2, \cdots, \mathrm{d}x_1\mathrm{d}x_t, \mathrm{d}x_2\mathrm{d}x_3, \cdots, \mathrm{d}x_{t-1}\mathrm{d}x_t)'$$

$$Z = \mathrm{d}X^3 = (\mathrm{d}x_1^3, \cdots, \mathrm{d}x_t^3, \mathrm{d}x_1^2\mathrm{d}x_2, \cdots, \mathrm{d}x_t^2\mathrm{d}x_{t-1}, \mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3, \cdots, \mathrm{d}x_{t-2}\mathrm{d}x_{t-1}\mathrm{d}x_t)' \tag{3-5-5}$$

由(3-5-1)式知，Y 和 Z 相当于附加参数，即在线性观测方程中增加了两组附加参数 Y 和 Z. 由于 Y 和 Z 都是 dX 的非线性函数，使得参数之间不独立. 故由(3-5-2)式不可能同时解出 dX、Y、Z. 由测量平差知，既然参数 dX 与附加参数 Y、Z 之间相关，那么 dX 与 Y、Z 之间就存在限制条件. 其限制条件的个数 S 等于参数总数减去必要观测值的个数，即 $S =$

$(t+\alpha+\beta)-t=\alpha+\beta$. 于是可得附有限制条件的间接平差模型:

$$
\begin{cases}
V = B\mathrm{d}x + \dfrac{1}{2}CY + \dfrac{1}{6}DZ + l \\
y_1 - \mathrm{d}x_1^2 = 0 \\
\vdots \\
y_t - \mathrm{d}x_t^2 = 0 \\
y_{t+1} - \mathrm{d}x_1\mathrm{d}x_2 = 0 \\
\vdots \\
y_\alpha - \mathrm{d}x_{t-1}\mathrm{d}x_t = 0 \\
Z_1 - \mathrm{d}x_1^3 = 0 \\
\vdots \\
Z_t - \mathrm{d}x_t^3 = 0 \\
Z_{t+1} - \mathrm{d}x_1^2\mathrm{d}x_2 = 0 \\
\vdots \\
Z_\beta - \mathrm{d}x_{t-2}\mathrm{d}x_{t-1}\mathrm{d}x_t = 0
\end{cases}
\tag{3-5-6}
$$

(3-5-6)式中的限制条件也是非线性方程. 仍需展为泰勒级数. 由于 $\mathrm{d}X$ 是 X_0 的改正数, 已是微小量, 故将条件方程展为泰勒级数时, 只需取至一次项即可. 为此, 令

$$
\mathrm{d}X = \mathrm{d}X_0 + \delta \tag{3-5-7}
$$

将(3-5-6)式中的限制条件展开, 得

$$
Y = Y_0 + \mathrm{d}Y = \begin{pmatrix} \mathrm{d}x_{01}^2 \\ \vdots \\ \mathrm{d}x_{0t}^2 \\ \mathrm{d}x_{01}\mathrm{d}x_{02} \\ \vdots \\ \mathrm{d}x_{0t-1}\mathrm{d}x_{0t} \end{pmatrix} + \begin{pmatrix} 2\mathrm{d}x_{01}\delta_1 \\ \vdots \\ 2\mathrm{d}x_{0t}\delta_2 \\ \mathrm{d}x_{02}\delta_1 + \mathrm{d}x_{01}\delta_2 \\ \vdots \\ \mathrm{d}x_{0t}\delta_{t-1} + \mathrm{d}x_{0t-1}\delta_t \end{pmatrix} \tag{3-5-8}
$$

$$
Z = Z_0 + \mathrm{d}Z = \begin{pmatrix} \mathrm{d}x_{01}^3 \\ \vdots \\ \mathrm{d}x_{0t}^3 \\ \mathrm{d}x_{01}^2\mathrm{d}x_{02} \\ \vdots \\ \mathrm{d}x_{0t}^2\mathrm{d}x_{t-1}^0 \\ \mathrm{d}x_{01}\mathrm{d}x_{02}\mathrm{d}x_{03} \\ \vdots \\ \mathrm{d}x_{0t-2}\mathrm{d}x_{0t-1}\mathrm{d}x_{0t} \end{pmatrix} + \begin{pmatrix} 3\mathrm{d}x_{01}^3\delta_1 \\ \vdots \\ 3\mathrm{d}x_{0t}^3\delta_t \\ 2\mathrm{d}x_{01}\mathrm{d}x_{02}\delta_1 + \mathrm{d}x_{01}^2\delta_2 \\ \vdots \\ 2\mathrm{d}x_{0t}\mathrm{d}x_{0t-1}\delta_t + \mathrm{d}x_{0t}^2\mathrm{d}x_{t-1} \\ \mathrm{d}x_{02}\mathrm{d}x_{03}\delta_1 + \mathrm{d}x_{01}\mathrm{d}x_{03}\delta_2 + \mathrm{d}x_{01}\mathrm{d}x_{03}\delta_3 \\ \vdots \\ \mathrm{d}x_{0t-1}\mathrm{d}x_{0t}\delta_1 + \mathrm{d}x_{0t-2}\mathrm{d}x_{0t}\delta_{t-1} + \mathrm{d}x_{0t-2}\mathrm{d}x_{0t-1}\delta_t \end{pmatrix} \tag{3-5-9}
$$

将(3-5-8)式和(3-5-9)式代入(3-5-6)式的第一式, 消去附加参数 Y 和 Z, 得新的误差方程为

$$V = B\mathrm{d}x_0 + B\delta + \frac{1}{2}Cy_0 + \bar{C}\delta + \frac{1}{6}DX_0 + \bar{D}\delta + l \tag{3-5-10}$$

经推导，可得

$$\bar{C} = \begin{bmatrix} \sum_{i=1}^{t}\mathrm{d}x_{0i}\frac{\partial^2 f_1}{\partial x_1 \partial x_i} & \sum_{i=1}^{t}\mathrm{d}x_{0i}\frac{\partial^2 f_1}{\partial x_2 \partial x_i} & \cdots & \sum_{i=1}^{t}\mathrm{d}x_{0i}\frac{\partial^2 f_1}{\partial x_t \partial x_i} \\ \sum_{i=1}^{t}\mathrm{d}x_{0i}\frac{\partial^2 f_2}{\partial x_1 \partial x_i} & \sum_{i=1}^{t}\mathrm{d}x_{0i}\frac{\partial^2 f_2}{\partial x_2 \partial x_i} & \cdots & \sum_{i=1}^{t}\mathrm{d}x_{0i}\frac{\partial^2 f_2}{\partial x_t \partial x_i} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{t}\mathrm{d}x_{0i}\frac{\partial^2 f_n}{\partial x_1 \partial x_i} & \sum_{i=1}^{t}\mathrm{d}x_{0i}\frac{\partial^2 f_n}{\partial x_2 \partial x_i} & \cdots & \sum_{i=1}^{t}\mathrm{d}x_{0i}\frac{\partial^2 f_n}{\partial x_t \partial x_i} \end{bmatrix} \tag{3-5-11}$$

$$\bar{D} = \frac{1}{2}\begin{bmatrix} \bar{d}_{11} & \bar{d}_{12} & \cdots & \bar{d}_{1k} & \cdots & \bar{d}_{1t} \\ \bar{d}_{21} & \bar{d}_{22} & \cdots & \bar{d}_{2k} & \cdots & \bar{d}_{2t} \\ \vdots & \vdots & & \vdots & & \vdots \\ \bar{d}_{n1} & \bar{d}_{n2} & \cdots & \bar{d}_{nk} & \cdots & \bar{d}_{nt} \end{bmatrix} \tag{3-5-12}$$

其中：

$$\bar{d}_{11} = \sum_{i=1}^{t}\mathrm{d}x_{0i}^2\frac{\partial^3 f_1}{\partial x_i^2 \partial x_1} + 2\mathrm{d}x_{01}^2\sum_{i=2}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_1}{\partial x_1^2 \partial x_i} + 2\sum_{j=2}^{t-1}\mathrm{d}x_{0j}\sum_{i\neq j+1}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_1}{\partial x_1 \partial x_j \partial x_i}$$

$$\bar{d}_{1k} = \sum_{i=1}^{t}\mathrm{d}x_{0i}^2\frac{\partial^3 f_1}{\partial x_i^2 \partial x_k} + 2\mathrm{d}x_{0k}\sum_{\substack{i=1\\ i\neq k}}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_1}{\partial x_k^2 \partial x_i} + 2\mathrm{d}x_{01}\sum_{\substack{i=2\\ i\neq k}}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_1}{\partial x_1 \partial x_k \partial x_i}$$
$$+ 2\sum_{\substack{j=2\\ j\neq k}}^{t-1}\mathrm{d}x_{0j}\sum_{\substack{i=j+1\\ i\neq k}}^{t-1}\mathrm{d}x_{0i}\frac{\partial^3 f_1}{\partial x_j \partial x_k \partial x_i}$$

$$\bar{d}_{1t} = \sum_{i=1}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_1}{\partial x_i^2 \partial x_t} + 2\mathrm{d}x_{0t}\sum_{i=1}^{t-1}\mathrm{d}x_{0i}\frac{\partial^3 f_1}{\partial x_k^2 \partial x_t} + 2\mathrm{d}x_{01}\sum_{i=2}^{t-1}\mathrm{d}x_{0i}\frac{\partial^3 f_1}{\partial x_1 \partial x_i \partial x_t}$$
$$+ 2\sum_{j=2}^{t-2}\mathrm{d}x_{0j}\sum_{i=j+1}^{t-1}\mathrm{d}x_{0i}\frac{\partial^3 f_1}{\partial x_j \partial x_i \partial x_t}$$

$$\bar{d}_{21} = \sum_{i=1}^{t}\mathrm{d}x_{0i}^2\frac{\partial^3 f_2}{\partial x_i^2 \partial x_1} + 2\mathrm{d}x_{01}\sum_{i=2}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_2}{\partial x_1^2 \partial x_i} + 2\sum_{j=2}^{t-1}\mathrm{d}x_{0j}\sum_{i=j+1}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_2}{\partial x_1 \partial x_j \partial x_i}$$

$$\bar{d}_{2k} = \sum_{i=1}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_2}{\partial x_i^2 \partial x_k} + 2\mathrm{d}x_{0k}\sum_{\substack{i=1\\ i\neq k}}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_2}{\partial x_k^2 \partial x_i} + 2\mathrm{d}x_{01}\sum_{\substack{i=2\\ i\neq k}}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_2}{\partial x_1 \partial x_k \partial x_i}$$
$$+ 2\sum_{\substack{j=2\\ j\neq k}}^{t-1}\mathrm{d}x_{0j}\sum_{\substack{i=j+1\\ i\neq k}}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_2}{\partial x_j \partial x_k \partial x_i}$$

$$\bar{d}_{2t} = \sum_{i=1}^{t}\mathrm{d}x_{0i}^2\frac{\partial^3 f_2}{\partial x_i^2 \partial x_t} + 2\mathrm{d}x_{0t}\sum_{i=1}^{t-1}\mathrm{d}x_{0i}\frac{\partial^3 f_2}{\partial x_t^2 \partial x_i} + 2\mathrm{d}x_{01}\sum_{i=2}^{t-1}\mathrm{d}x_{0i}\frac{\partial^3 f_2}{\partial x_1 \partial x_i \partial x_t}$$
$$+ 2\sum_{j=2}^{t-2}\mathrm{d}x_{0j}\sum_{i=j+1}^{t-1}\mathrm{d}x_{0i}\frac{\partial^3 f_2}{\partial x_j \partial x_i \partial x_t}$$

$$\bar{d}_{n1} = \sum_{i=1}^{t}\mathrm{d}x_{0i}^2\frac{\partial^3 f_n}{\partial x_i^2 \partial x_1} + 2\mathrm{d}x_{01}\sum_{i=2}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_n}{\partial x_1^2 \partial x_i} + 2\sum_{j=2}^{t-1}\mathrm{d}x_{0j}\sum_{i=j+1}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_n}{\partial x_1 \partial x_j \partial x_i}$$

$$\bar{d}_{nk} = \sum_{i=1}^{t}\mathrm{d}x_{0i}^2\frac{\partial^3 f_n}{\partial x_i^2 \partial x_k} + 2\mathrm{d}x_{0k}\sum_{\substack{i=1\\ i\neq k}}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_n}{\partial x_k^2 \partial x_i} + 2\mathrm{d}x_{01}\sum_{\substack{i=2\\ i\neq n}}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_n}{\partial x_1 \partial x_k \partial x_i}$$

$$+2\sum_{\substack{j=2\\j\neq n}}^{t-1}\mathrm{d}x_{0j}\sum_{\substack{i=j+1\\i\neq n}}^{t}\mathrm{d}x_{0i}\frac{\partial^3 f_n}{\partial x_j\partial x_k\partial x_i}$$

$$\bar{d}_{nt}=\sum_{i=1}^{t}\mathrm{d}x_{0i}^2\frac{\partial^3 f_n}{\partial x_i^2\partial x_t}+2\mathrm{d}x_{0t}\sum_{i=1}^{t-1}\mathrm{d}x_{0i}\frac{\partial^3 f_n}{\partial x_i^2\partial x_t}+2\mathrm{d}x_{01}\sum_{i=2}^{t-1}\mathrm{d}x_{0i}\frac{\partial^3 f_n}{\partial x_1\partial x_i\partial x_t}$$

$$+2\sum_{j=2}^{t-2}\mathrm{d}x_{0j}\sum_{i=j+1}^{t-1}\mathrm{d}x_{0i}\frac{\partial^3 f_n}{\partial x_j\partial x_i\partial x_t}$$

以上各式中,k 为$\bar{D}$ 矩阵的第k 列,且$2\leqslant k\leqslant t-1$.

令

$$\left.\begin{aligned}&\bar{B}=B+\bar{C}+\bar{D}\\&\bar{l}=B\mathrm{d}X_0+\frac{1}{2}CY_0+\frac{1}{6}DZ_0+l\end{aligned}\right\}\tag{3-5-13}$$

则(3-5-10)式可写为

$$V=\bar{B}\delta+\bar{l}\tag{3-5-14}$$

(3-5-14)式是我们所熟悉的一般线性模型的误差方程,根据最小二乘原理可得法方程

$$\bar{B}'\bar{B}\delta+\bar{B}'\bar{l}=0\tag{3-5-15}$$

二、解的惟一性定理

定理 3-5-1 当未知参数的个数等于必要观测数 t 并独立时,(3-5-15)式的解存在且惟一.

证明: 当未知数的个数等于必要观测数 t 并独立时,有 $\mathrm{rk}(B)=t$,现设 $\mathrm{rk}(\bar{C})=t_C$,$\mathrm{rk}(\bar{D})=t_D$,显然 t_C、t_D 都不大于t,即 $t_C\leqslant t$,$t_D\leqslant t$.于是

$$\mathrm{rk}(\bar{B})=\mathrm{rk}(B+\bar{C}+\bar{D})\leqslant\mathrm{rk}(B)+\mathrm{rk}(\bar{C})+\mathrm{rk}(\bar{D})=t+t_C+t_D$$

由于 $t_C\neq0$, $t_D\neq0$,所以

$$\mathrm{rk}(\bar{B})\geqslant\mathrm{rk}(B)=t\tag{3-5-16}$$

又因为 $\bar{B}$ 是$n\times t$ 的矩阵,而 $n>t$,所以 $\bar{B}$ 的秩不大于t,即

$$\mathrm{rk}(\bar{B})\leqslant t\tag{3-5-17}$$

综合(3-5-16)式和(3-5-17)式得

$$\mathrm{rk}(\bar{B})=t$$

所以有 $\mathrm{rk}(\bar{B}'\bar{B})=t$,即 $\bar{B}'\bar{B}$ 为满秩方阵,于是(3-5-14)式有惟一解.

于是

$$\delta=-(\bar{B}'\bar{B})^{-1}\bar{B}'\bar{l}\tag{3-5-18}$$

$$\hat{X}=X_0+\mathrm{d}X_0+\delta\tag{3-5-19}$$

(3-5-19)式就是非线性模型(2-1-4)式的直接解算结果.

三、直接解算步骤

由(3-5-17)式知,δ 的解算与传统方法相似.所不同的就是在按(3-5-17)式求 δ 之前,要先用参数的改正数 $\mathrm{d}X$ 的近似值 $\mathrm{d}X_0$ 计算出 Y_0, Z_0 以及矩阵 C, D, $\bar{C}$ 和$\bar{D}$.具体解算步骤为:

(1) 按传统方法确定参数 X 的近似值X_0.

(2) 对非线性模型(2-1-4)式中的非线性函数 $f(X)$ 分别求一、二、三阶偏导数. 并用 X_0 计算 B、C 和 D 矩阵.

(3) 选定适当小的向量作为 dX_0, 或按传统的线性近似方法求 dX_0, 即 $dX_0 = -(B'B)^{-1}B'l$

(4) 根据(3-5-8)式和(3-5-9)式计算 Y_0 和 Z_0.

(5) 根据(3-5-11)式和(3-5-12)式计算 $\bar{C}$ 和 $\bar{D}$.

(6) 根据(3-5-13)式计算 $\bar{B}$ 和 $\bar{l}$, 然后根据(3-5-18)式计算 δ, 最后根据(3-5-19)式计算 $\hat{X}$.

例 3-5-1 设 $X_0 = (5.4 \quad -0.3)'$, $dX_0 = (-0.005859 \quad 0.049949)'$, 用直接解法解算例 2-1-1 中的非线性模型.

按以上算法编写了计算机程序, 输出结果为

$$\delta = (0.0256889296 \quad -0.005056793025)'$$

$$\hat{X} = (5.41882993 \quad -0.255107793)', \ \|\Delta X\| = 0.0015$$

$$R = -40.62921193$$

若 X_0 不变, $dX_0 = (-0.006 \quad 0.045)'$, 则输出结果为

$$\delta = (0.0266988236 \quad -0.0003768357749)'$$

$$\hat{X} = (5.420698824 \quad -0.2553768358)', \ \|\Delta X\| = 0.0012$$

$$R = -40.63406453$$

由此可见, 直接算法强烈地依赖 dX_0 的值. 只有在 dX_0 较精确时, 才能得到较精确的结果. 由于 dX_0 在求解前是无法知道的, 因此无法得到参数的高精度估值, $V'V$ 也无法达到最小.

第六节　非线性最小二乘估计的其他算法

前面介绍的各种算法, 都需要求导计算. 当 $f(X)$ 比较复杂时, 求导很困难. 尤其是当 $f(X)$ 不可导时, 前面介绍的各种方法均不能应用. 为此, 下面集中介绍几种无需求导计算的直接搜索算法.

一、单纯形法

单纯形法是按给定方案逐步搜索, 直接比较目标函数的大小来寻求其极小值点的一种方法[56] (林洪桦, 1995). 单纯形是指以 t 维空间中的 $t+1$ 个点作为顶点所构成的超多面体. 例如, 二维空间中的三角形, 三维空间中的四面体等均为单纯形. 单纯形法的基本原理如下:

对于非线性模型中的 t 个待估参数, 按一定的规则选取 $t+1$ 组近似值, 构成初始单纯形. 用此 $t+1$ 组近似值计算 $t+1$ 个目标函数值并比较其大小. 找出最大的目标函数值, 并剔除相应的那组参数近似值. 然后按一定的换点规则换入新的一组参数近似值. 用这组新的近似值与其余的 t 组近似值构成新的单纯形. 如此反复计算、比较、剔除, 直至最小的目标函数值以给定的精度逼近其极小值为止. 这时与目标函数最小值相对应的那组参数近似值即为参数的非线性最小二乘估计.

实现该原理首先要确定初始单纯形，然后要制定换点规则．这两个问题可采用如下方法解决：

1．选取初始单纯形

对于非线性模型中的 t 个待估参数，通常选取 $t+1$ 组参数近似值，构成具有 $t+1$ 个顶点的正则单纯形（棱长相等的单纯形称为正则单纯形）．一般先任选一组参数近似值 X_0 作为顶点，其余 t 个顶点 $X_1, X_2, \cdots, X_t$ 按等棱长 l 选取，即

$$\left.\begin{array}{l} X_0=(x_{01} \quad x_{02} \quad \cdots \quad x_{0t})' \\ X_1=(x_{01}+c \quad x_{02}+d \quad \cdots \quad x_{0t}+d)' \\ X_2=(x_{01}+d \quad x_{02}+c \quad \cdots \quad x_{0t}+d)' \\ \qquad\vdots \\ X_t=(x_{01}+d \quad x_{02}+d \quad \cdots \quad x_{0t}+c)' \end{array}\right\} \tag{3-6-1}$$

式中：

$$c=\frac{\sqrt{t+1}+t-1}{\sqrt{2}t}l, \quad d=\frac{\sqrt{t+1}-1}{\sqrt{2}t}l \tag{3-6-2}$$

于是，初始单纯形的任意两个顶点间的距离为 l，即

$$\| X_0-X_i \| = \sqrt{c^2+(t-1)d^2}=l$$

$$\| X_i-X_j \| = \sqrt{2}(c-d)=l; \quad i,j=1,2,\cdots,t; \quad i\neq j.$$

例如，对于例 2-1-1 中的非线性模型，取 $l=5$，用第一个观测值和第二个观测值可算得 $X_0'=(5.428 \quad -0.256)$，由(3-6-2)式得 $c=4.829629131$，$d=1.294095226$．于是初始单纯形的三个顶点为：

$$(5.428 \quad -0.256),(10.256 \quad 1.038),(6.722 \quad 4.574)$$

初始单纯形于图 3-6-1 所示．

当然，初始单纯形也可用其他方法选取．但不论用什么方法选取，都应该保证初始单纯形的 $t+1$ 个顶点中，除 X_0 外，其余 t 个顶点与 X_0 之差所构成的 t 个向量

$$X_1-X_0, \ X_2-X_0, \ \cdots, \ X_t-X_0$$

线性无关，即 3 个顶点不共线，4 个顶点不共面等．

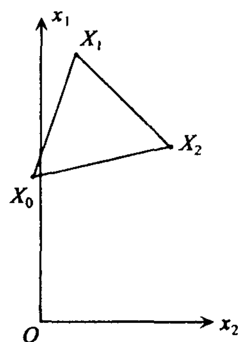


图 3-6-1

2．换点规则

剔除目标函数值最大的顶点后，应换入新的顶点．新顶点的换入需遵循一定的规则．换点规则分为反射、伸长、缩短和缩小四种情况．为了叙述方便，记最大目标函数值为 R_U，相应的顶点为 X_U；最小目标函数值为 R_L，相应的顶点为 X_L．下面结合初始单纯形图 3-6-1 来分别介绍上述 4 种规则．

由例 3-2-1 知，目标函数为

$$\begin{aligned} R(X) &= f'(X)f(X)-2f'(X)L \\ &= x_1^2\sum_{i=1}^{5}\mathrm{e}^{2ix_2}-2x_1\sum_{i=1}^{5}L_i\mathrm{e}^{ix_2} \end{aligned} \tag{3-6-3}$$

(1)反射

将初始单纯形的三个顶点分别代入目标函数(3-6-3)式,可得:

$$R(X_0)=-40.63548307,\ R(X_1)=3\ 866\ 953.823,\ R(X_2)=3.29666\times10^{21}$$

剔除目标函数值最大的顶点 X_U(即 X_2). 计算其余各顶点的中心点 X_C, 即

$$X_C=\left(\frac{x_{01}+x_{11}}{2}\quad\frac{x_{02}+x_{12}}{2}\right)'$$

再计算目标函数值最大的顶点 X_2 关于该中心点 X_C 的反射点 X_R, 即

$$X_R=(2x_{c1}-x_{21}\quad 2x_{c2}-x_{22})$$

图 3-6-2 表示该例的反射. 用新顶点 X_R 取代 X_U, 称为反射.

(2)伸长

将反射点 X_R 代入(3-6-3)式,计算反射点的目标函数值 R_R. 若 $R_R<R_U$, 表明反射方向能使目标函数下降. 沿此方向伸长取新点,称为伸长. 伸长点用 X_E 表示,则其坐标为:

$$x_{Ei}=(1+\eta)x_{Ri}-\eta x_{Ci}\quad(i=1,2,\cdots,t)\tag{3-6-4}$$

式中: η 称为伸长因子,一般取 $\eta=2$.

伸长后,计算伸长点 X_E 的目标函数值 R_E. 若 $R_E<R_U$, 则用 X_E 取代 X_U, 构成新的单纯形. 图 3-6-3 表示该例的伸长.

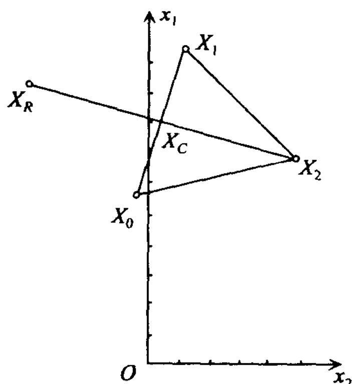


图 3-6-2

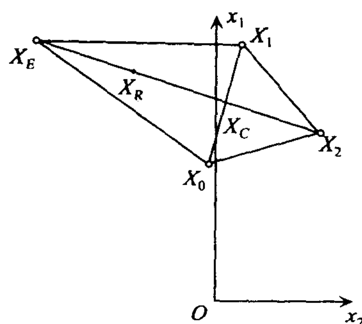


图 3-6-3

(3)缩短

反射后,若 $R_R>R_U$, 表明反射点仍是最坏点,即反射方向不是目标函数的下降方向. 这时可自反射点 X_R 向中心点 X_C 缩短距离,取新点 X_S. 自反射点 X_R 向中心点 X_C 缩短距离,称为缩短. 缩短点 X_S 的坐标为:

$$x_{Si}=(1-\beta)x_{Ci}+\beta x_{Ri}\tag{3-6-5}$$

式中: β 为缩短因子,一般取 $\beta=0.5$.

缩短后,计算缩短点 X_S 的目标函数值 R_S. 若 $R_S<R_U$, 则用 X_S 取代 X_U, 构成新的单纯形. 图 3-6-4 表示该例的缩短.

(4)缩小

当反射无效,即 R_R、R_E 和 R_S 均不小于 R_U 时,可将整个单纯形向目标函数值最小点 X_L 处缩小. 通常将单纯形缩小一半,即

$$x_{ij}=\frac{x_{ij}+x_{Lj}}{2}\quad(i=0,1,\cdots,t;\ j=1,2,\cdots,t)\tag{3-6-6}$$

图 3-6-5 表示该例的缩小.

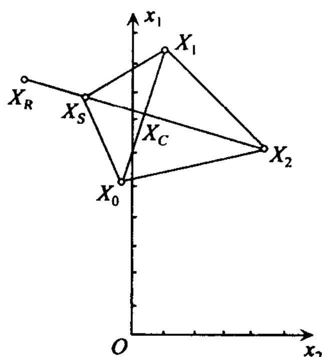


图 3-6-4

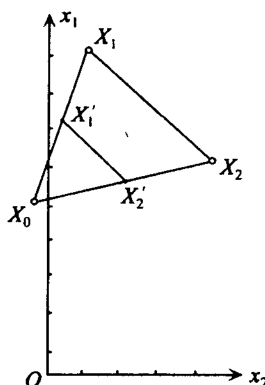


图 3-6-5

按上述换点规则逐步变换单纯形,就可以使最终的单纯形顶点的目标函数最小值以给定的精度逼近其极小值.搜索终止条件为

$$\sqrt{\frac{1}{t+1}\sum_{i=0}^{t}(R(X_i)-R_L)^2}<\delta\tag{3-6-7}$$

式中:$\delta>0$ 为给定的误差限.

根据此换点规则,可总结出单纯形法的计算步骤如下:

①选取初始单纯形.

②计算各顶点的目标函数值 $R(X_i)$.若 $R(X_i)$满足(3-6-7)式,则转⑧.

③比较各顶点目标函数值的大小,找出最大的目标函数值 R_U 和最小的目标函数值 R_L.将目标函数值最大的顶点和目标函数值最小的顶点分别标定为 X_U 和 X_L,并确定中心点 X_C.

④确定反射点 X_R,并计算反射点 X_R 的目标函数值 R_R.将 R_R 与 R_U 和 R_L 进行比较:

(a)若 $R_R\leqslant R_L$,则转⑤;

(b)若 $R_R\geqslant R_U$,则转⑥;

(c)若 $R_L<R_R<R_U$,则以 X_R 代替 X_U 构成新的单纯形,转②.

⑤按(3-6-4)式计算伸长点 X_E,并计算伸长点 X_E 的目标函数值 R_E.比较 R_E 与 R_L:

(a)若 $R_E<R_L$,则以 X_E 代替 X_U 构成新的单纯形,转②;

(b)若 $R_E\geqslant R_L$,则仍以 X_R 代替 X_U 构成新的单纯形,转②.

⑥按(3-6-5)式计算缩短点 X_S,并计算缩短点 X_S 的目标函数值 R_S.比较 R_S 与 R_U:

(a)若 $R_S\geqslant R_U$,则转⑦;

(b)若 $R_S<R_L$,则以 X_S 代替 X_U 构成新的单纯形,转②.

⑦按(3-6-6)式缩小单纯形后,转②.

⑧停止搜索,输出 X_L.

根据此计算步骤,我们编写程序并计算了例 2-1-1 中非线性模型参数的非线性最小二

乘估计.初始单纯形如上所述,经 70 次搜索,得

$X=(5.427324646 \quad -0.255898000)$, $\|\Delta X\|=0.0074$, $R(X)=-40.63548408$.

通过实际计算,我们体会到:单纯形法虽然具有不需求导计算的优点,但若按以上换点规则搜索,只能得到局部最小.而且顶点距离 l、伸长因子 η 以及缩短因子 β 等都对搜索有较大的影响.因此,单纯形法并不是很好的算法.除非研究新的换点规则.

二、模拟退火算法

模拟退火算法(Simulated Annealing Algorithm, SAA)是一种求目标函数极小值的全局优化算法.其基本思想来源于统计物理学[57](李士勇,1998).

当金属在高温下熔化时,所有原子都处于高能的自由运动状态.随着温度的降低,原子的自由运动减弱,物体的能量随之降低.只要在凝结温度附近使温度下降足够慢,原子排列就越来越有规则,从而形成结晶.这一过程称为退火过程.这一过程可用来模拟多元目标函数的优化过程.用物理上的退火过程来模拟多元目标函数的优化过程,称为模拟退火算法.

模拟退火算法可概述为:

① 选取初始点 X_0 和初始温度 T,并令 $k=1$.

② 计算初始点 X_0 的目标函数值 $R(X_0)$.

③ 若 $T=0$,则转⑥.

④ 对 X_0 进行随机扰动,得到新点 X_1.

⑤ 计算新点 X_1 的目标函数值 $R(X_1)$和目标函数之差 $\Delta R=R(X_1)-R(X_0)$:

(a)若 $\Delta R<0$,则降温:

$$T=\frac{T}{\lg(10+\sqrt{k})} \tag{3-6-8}$$

且令 $X_0=X_1$, $R(X_0)=R(X_1)$, $X=X_1$, $k=k+1$,转③.

(b)若 $\Delta R>0$,则产生一个[0,1]上的随机数 A,并按(3-6-9)式计算接受非优点 X_1 的概率:

$$P=\exp\left(-\frac{\Delta R}{T}\right) \tag{3-6-9}$$

然后比较 A 与 P:

b.1 若 $P>A$,则接受非优点,即 $X_0=X_1$,转③.

b.2 转③.

⑥ 输出 X_0.

按此算法,有一定的概率跳出局部最优解.从而有可能找到全局最优解.由(3-6-7)式和(3-6-8)式知,接受非优点的概率由 ΔR 和温度 T 确定.当温度 T 一定时,ΔR 越大,接受非优点的概率就越小;ΔR 越小,接受非优点的概率就越大.当 $\Delta R=0$ 时,只要 T 不等于 0,则接受非优点的概率为 1;因此,为了在 ΔR 较小时不接受非优点,可对 ΔR 进行限制.比如,当 $\Delta R<1$ 时,令 $\Delta R=1$.另外,当 ΔR 一定时,由(3-6-8)式知,搜索开始时,由于温度较高,接受非优点的概率较大.随着最优解的不断逼近,温度逐渐下降,接受非优点的概率也逐渐减小.

模拟退火算法虽然可得到全局最优解,但是最优解的搜索是在整个定义域中反复随机

搜索,搜索时间很长.

应用模拟退火算法,我们编写程序并计算了例 2-1-1 中非线性模型参数的非线性最小二乘估计. 经很长时间的搜索,温度下降到 0,得

$X=(5.423806309 \quad -0.255871001)$, $\| \Delta X \| = 0.0040$, $R(X) = -40.635488853$.

三、遗传算法

遗传算法(Genetic Algorithm, AG)是建立在自然选择和自然遗传学机理之基础上的迭代自适应概率性搜索算法[57](李士勇,1998). 该算法通过编码、初始群体生成、适应度评价检测、选择、交叉、变异等过程模拟生物进化,实现模型自学习与优化的目的[58](陈国良、王熙法、庄镇泉等,1996). 这一算法最先由美国密执安大学的 J. H. Holland 教授于 1975 年提出. 1975 年 J. H. Holland 出版了遗传算法历史上的经典著作《自然和人工系统中的适应性》,该著作系统地阐述了遗传算法的基本理论和方法,并提出了模式定理(Schemata Theorem). 证明在遗传算子选择、交叉和变异的作用下,具有低阶、短定义距以及平均适应度高于群体平均适应度的模式在子代中将以指数级增长. 这里的模式是某一类字符串,其某些位置有相似性. 同年,DeJong 完成了他的博士论文《遗传自适应系统的行为分析》,将 Holland 的模式理论与自己的计算试验结合起来,进一步完善了选择、交叉和变异操作,提出了一些新的遗传操作技术.

进入 20 世纪 80 年代后,遗传算法得到了迅速发展,不仅理论研究十分活跃,而且在越来越多的应用领域中得到应用. 1983 年,Holland 的学生 Goldberg 将遗传算法应用于管道煤气系统的优化,很好地解决了这一非常复杂的问题. 1989 年,Goldberg 出版了《搜索、优化和机器学习中的遗传算法》一书,这本书可能是遗传算法领域被引用次数最多的书,为这一领域奠定了坚实的科学基础.

遗传算法的一般过程由初始化、选择、交叉和突然变异四个部分组成. 其结构见图 3-6-6[57](李士勇,1998).

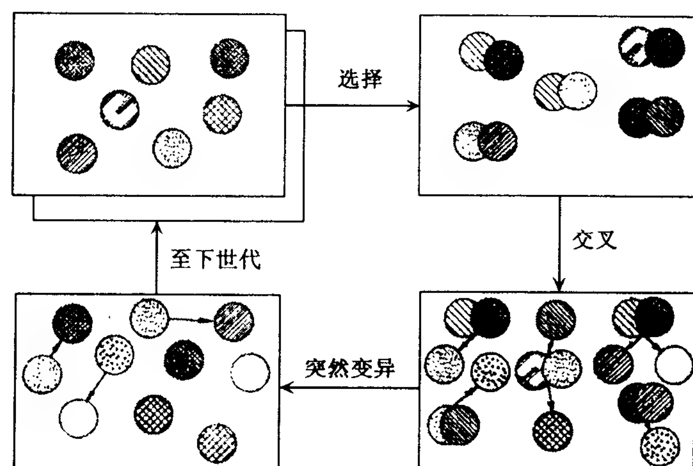


图 3-6-6

用遗传算法求最优解的步骤如下：

① 编码：用一定比特的0、1二进制数对自变量进行编码，形成基因码链.每一个码链代表一个个体.

② 产生群体：令 $t=0$，随机产生 n 个个体，形成群体 $P(t)$.该群体代表优化问题的一些可能解的集合.遗传算法就是从这些群体出发，模拟生物进化过程，存优汰劣，从而选择出优秀的群体和个体.

③ 评价：按编码规则，将群体 $P(t)$ 中的每一个个体的基因所对应的自变量值 X_i 代入目标函数，算出各个个体的目标函数值 $R(X_i)$. $R(X_i)$ 越小，说明该个体的适应度越高，更适合生存环境.如果最优个体的适应度达到某一阈值或最优个体的适应度和平均适应度不再下降，则转⑧.

④ 选择：按一定的概率从群体 $P(t)$ 中选出 m 个个体，作为双亲繁殖后代，产生新的个体加入下一个群体 $P(t+1)$ 中.

⑤ 交叉：对于选中的用于繁殖的每一对个体，随机地选择同一整数 n，将双亲的基因链在此位置相互交换.

⑥ 变异：以一定的概率 p 从群体 $P(t+1)$ 中随机选取若干个体.对于选中的个体，随机选取某一位进行取反运算，即由 $1\to0$ 或由 $0\to1$.变异模拟了生物进化过程中的基因突变现象.

⑦ 转③.

⑧ 输出最优解.

遗传算法具有许多独特的优点：

1.不必非常明确地描述问题的全部特征，其通用性和稳健性强，能很快适应问题和环境的变化；对领域知识依赖程度低，不受搜索空间限制性假设的约束，不必要求连续性、可导或单峰等.

2.从多点进行搜索，如同在搜索空间上覆盖的一张网，搜索的全局性强，不易陷入局部最优；具有并行性，非常适合于并行计算.

在遗传算法的研究中，主要有下列三类研究方向：

1.研究遗传算法本身的理论基础.

2.用遗传算法作为工具解决工程问题，主要是进行优化，关心的是是否能在传统方法上有所提高.

3.用遗传算法研究演化现象，一般涉及到人工生命和复杂性科学领域.

第四章 非线性最小二乘估计的统计性质与精度评定

第一节 非线性最小二乘估计的统计性质

在线性模型 $L=BX+\Delta$ 中,当 Δ 服从正态分布时,最小二乘估计 $\hat{X}$ 是参数 X 的最优无偏估计,并且有 $\mathrm{E}(V)=0$, $\mathrm{Var}(\hat{X})=\sigma^2(B'B)^{-1}$, $\mathrm{tr}\,(\mathrm{Var}(\hat{X}))=\min$, $E(V'V)=(n-t)\sigma^2$, $\mathrm{Var}(V)=\sigma^2(Q_{LL}-B(B'B)^{-1}B')$. 那么,在非线性模型 $L=f(X)+\Delta$ 中,当 Δ 仍服从正态分布时,$\hat{X}$、V 是否还有这些优良统计性质呢? 下面就来讨论这个问题[24](韦博成,1989).

一、正则条件

要研究非线性最小二乘估计的统计性质,需要相当繁复的正则条件. 作为在测量平差中的应用,我们在假定 Δ 为同精度独立正态随机误差,即 $\Delta\sim N(0,\sigma^2I)$的前提下,把条件予以放宽.

条件(ⅰ) 在非线性模型(2-1-4)式中,假设非线性函数 $f(X)$是紧集 $\mathscr{X}$上关于 X 的连续函数. 随机误差 $\Delta\sim N(0,\sigma^2I)$, 参数的真值记为 $\tilde{X}$, 它是 $\mathscr{X}$的内点.

条件(ⅱ) 当 $n\to\infty$时,函数

$$\varphi_n(X,\tilde{X})=\frac{1}{n}\sum_{i=1}^{n}(f_i(X)-f_i(\tilde{X}))^2$$

在 $\mathscr{X}\times\mathscr{X}$上一致收敛到 $\varphi(X,\tilde{X})$, 且 $\varphi(X,\tilde{X})$在 X 上存在惟一的极小值点.

条件(ⅲ) $f_i(X)$在 $\mathscr{X}$上关于 X 存在一阶连续导数,且在 $\tilde{X}$ 的某个邻域内有

$$\lim_{n\to\infty}\frac{1}{n}B'(X)B(X)=\Omega(X) \tag{4-1-1}$$

式中:$\Omega(X)$为正定矩阵.

条件(ⅳ) $f_i(X)$在 $\mathscr{X}$上关于 X 存在二阶连续导数,且在 $\tilde{X}$ 的某个邻域内有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2 f_i}{\partial x_i\partial x_j}\right)^2=E_{ij}(X) \tag{4-1-2}$$

条件(ⅴ) 假定模型函数 $f_i(X)$在 $\mathscr{X}$上存在三阶连续导数,并且下列有关导数的极限存在

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}\frac{\partial^2 f_i}{\partial x_j\partial x_k}=E_{ijk}(X)<+\infty$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^3 f_i}{\partial x_i\partial x_j\partial x_k}\right)^2=G_{ijk}(X)<+\infty$$

二、非线性最小二乘估计的统计性质

设非线性模型如(2-1-4)式所示，即 $L=f(X)+\Delta$，由于最小二乘估计 $\hat{X}(L)$和 $\Delta=L-f(X)$都是随机变量 $L=(L_1\quad L_2\quad\cdots\quad L_n)'$的函数，故称其展开式为随机展开[24]（韦博成，1989）. 下面以 $\hat{X}$ 的随机展开为基础来讨论非线性最小二乘估计的统计性质.

1. 渐近性

定理 4-1-1 若非线性模型(2-1-4)式满足正则条件(ⅰ)~(ⅳ)，则参数 X 的非线性最小二乘估计 $\hat{X}$ 为渐近正态的，且以下关系式成立：

$$\sqrt{n}(\hat{X}-\tilde{X})\xrightarrow{L}N(0,\sigma^2\Omega^{-1}(\tilde{X}))\quad(n\to\infty)\tag{4-1-3}$$

式中，符号$\xrightarrow{L}$表示依概率收敛于

$$\hat{X}-\tilde{X}=\Delta X=(B'B)^{-1}B'\Delta+a_n\tag{4-1-4}$$

其中 a_n 满足条件$\sqrt{n}a_n\to 0(n\to\infty)$.

证明： 先证(4-1-4)式. 因为 $V'V=\|f(X)-L\|^2$ 在 $\hat{X}$ 处达到最小值. 因此有 $\left.\frac{\partial V'V}{\partial X}\right|_{X=\hat{X}}=0$. 根据正则条件(ⅲ)和(ⅳ)，$\left.\frac{\partial V'V}{\partial X}\right|_{X=\hat{X}}$ 可在$\tilde{X}$ 处展开为

$$\left(\left.\frac{\partial V'V}{\partial X}\right|_{X=\hat{X}}\right)'=\left(\left.\frac{\partial V'V}{\partial X}\right|_{X=\tilde{X}}\right)'+\frac{\partial V'V}{\partial X^2}\Delta X+\gamma\Delta X=0$$

其中 $\Delta X=\hat{X}-\tilde{X}$，当 $\Delta X\to 0$ 时，$\gamma\to 0$. 因为

$$\left.\frac{\partial V'V}{\partial X}\right|_{X=\tilde{X}}=2\tilde{V}'\tilde{B}$$

$$\left.\frac{\partial^2 V'V}{\partial X^2}\right|_{X=\tilde{X}}=2\tilde{B}'\tilde{B}+2\tilde{V}'\tilde{W}$$

所以有

$$\tilde{B}'\tilde{V}+\tilde{B}'\tilde{B}+[\tilde{V}][\tilde{W}]\Delta X+\gamma\Delta X=0$$

顾及 $\tilde{V}=-\Delta$，得

$$\Delta X=(\tilde{B}'\tilde{B})^{-1}(\tilde{B}'\Delta-[\Delta'][\tilde{W}]\Delta X-\gamma\Delta X)$$

于是 ΔX 可写为

$$\Delta X=(\tilde{B}'\tilde{B})^{-1}\tilde{B}'\Delta+a_n$$

其中：

$$\begin{aligned}a_n&=-(\tilde{B}'\tilde{B})^{-1}([\Delta'][\tilde{W}]\Delta X+\gamma\Delta X)\\&=-\left(\frac{1}{n}\tilde{B}'\tilde{B}\right)^{-1}\left(\frac{1}{n}[\Delta'][\tilde{W}]\Delta X+\frac{1}{n}\gamma\Delta X\right)\end{aligned}$$

而

$$\sqrt{n}a_n=-\left(\frac{1}{n}\tilde{B}'\tilde{B}\right)^{-1}\left(\frac{1}{\sqrt{n}}[\Delta'][\tilde{W}]\Delta X+\left(\frac{1}{\sqrt{n}}\gamma\Delta X\right)\right)$$

由于 $\tilde{W}$ 的元素 w_{ijk}是函数$f_i(X)$的二阶偏导数，故当 $n\to\infty$时，有

$$\frac{1}{\sqrt{n}}[\Delta'][\widetilde{W}]\Delta X \to 0 \tag{4-1-5}$$

$$\frac{1}{\sqrt{n}}\gamma\Delta X \to 0$$

所以$\sqrt{n}a_n \to 0$，于是(4-1-4)式得证.

下面再来证明(4-1-3)式.

将(4-1-4)式两边乘以$\sqrt{n}$，得

$$\sqrt{n}(\hat{X}-\widetilde{X}) = \left(\frac{1}{n}\widetilde{B}'\widetilde{B}\right)^{-1}\left(\frac{1}{\sqrt{n}}\widetilde{B}'\Delta + \sqrt{n}a_n\right) \tag{4-1-6}$$

对(4-1-6)式取期望，得

$$\mathrm{E}(\sqrt{n}(\hat{X}-\widetilde{X})) = \left(\frac{1}{n}\widetilde{B}'\widetilde{B}\right)^{-1}\left(\frac{1}{\sqrt{n}}\widetilde{B}'\mathrm{E}(\Delta) + \mathrm{E}\sqrt{n}a_n\right)$$

因为 $\mathrm{E}(\Delta)=0, \sqrt{n}a_n \to 0 (n\to\infty)$

所以当 $n\to\infty$ 时，有

$$\mathrm{E}(\sqrt{n}(\hat{X}-\overline{X})) = 0 \tag{4-1-7}$$

又当 $n\to\infty$ 时，(4-1-6)式可写为

$$\sqrt{n}(\hat{X}-\widetilde{X}) = \left(\frac{1}{n}\widetilde{B}'\widetilde{B}\right)^{-1}\left(\frac{1}{\sqrt{n}}\widetilde{B}'\right)\Delta \tag{4-1-8}$$

根据(4-1-8)式，应用协方差传播律，得

$$\mathrm{Var}(\sqrt{n}(\hat{X}-\widetilde{X})) = \left(\frac{1}{n}\widetilde{B}'\widetilde{B}\right)^{-1}\left(\frac{1}{\sqrt{n}}\widetilde{B}'\right)(\sigma^2 I)\left(\frac{1}{\sqrt{n}}\widetilde{B}\left(\frac{1}{n}\widetilde{B}'\widetilde{B}\right)^{-1}\right) = \sigma^2\left(\frac{1}{n}\widetilde{B}'\widetilde{B}\right)^{-1}$$

由条件(ⅳ)知

$$\left(\frac{1}{n}\widetilde{B}'\widetilde{B}\right)^{-1} = \Omega^{-1}(\widetilde{X})$$

所以，当 $n\to\infty$ 时，有

$$\sqrt{n}(\hat{X}-\widetilde{X}) \xrightarrow{L} N(0, \sigma^2\Omega^{-1}(\widetilde{X}))$$

于是(4-1-3)式得证.

定理 4-1-2 若非线性模型(2-1-4)式满足正则条件(ⅰ)~(ⅳ). 则估计量 $S^2 = \frac{V'V}{n-t}$ 满足以下关系：

$$S^2 \xrightarrow{L} \frac{\sigma^2}{n-t}\chi^2(n-t) \quad (n\to\infty) \tag{4-1-9}$$

$$S^2 = \frac{1}{n-t}\Delta' S_N \Delta + a_n \tag{4-1-10}$$

其中：$S_N = I - S_T$，$S_T - \widetilde{B}(\widetilde{B}'\widetilde{B})^{-1}\widetilde{B}'$，$a_n \to 0$，而且 S^2 与 $\widetilde{B}(\hat{X}-\widetilde{X})$渐近独立.

证明：欲证明(4-1-9)式和(4-1-10)式，只需证明：

$$V'V = \Delta' S_N \Delta + \beta_n, \quad \beta_n \to 0 \tag{4-1-11}$$

即可.

由条件(ⅲ)和(ⅳ)可知，$V'V$ 在真值$\widetilde{X}$ 处展为泰勒级数可取至二次项，即

$$V'V = \widetilde{V}'\widetilde{V} + 2\widetilde{V}'\widetilde{B}(\hat{X}-\widetilde{X}) + (\hat{X}-\widetilde{X})'\widetilde{B}'\widetilde{B}(\hat{X}-\widetilde{X})$$

$$+(\hat{X}-\widetilde{X})'[\widetilde{V}'][\widetilde{W}](\hat{X}-\widetilde{X})+\gamma\|\hat{X}-\widetilde{X}\|^2$$

由(4-1-4)式知

$$\widetilde{B}(\hat{X}-\widetilde{X})=S_T\Delta+\widetilde{B}a_n \tag{4-1-12}$$

顾及 $\widetilde{V}=-\Delta,\widetilde{V}'\widetilde{V}=\Delta'\Delta$，得

$$\begin{aligned}V'V&=\Delta'\Delta-2\Delta'S_T\Delta-2\Delta'\widetilde{B}a_n+(S_T\Delta+\widetilde{B}a_n)'(S_T\Delta+\widetilde{B}a_n)\\&\quad-(\Delta X)'[\Delta'][\widetilde{W}](\Delta X)+\gamma\|\Delta X\|^2\\&=\Delta'\Delta-\Delta'S_T\Delta+a_n'\widetilde{B}'\widetilde{B}a_n-(\Delta X)'[\Delta]'[\widetilde{W}](\Delta X)+\gamma\|\Delta X\|^2\\&=\Delta'S_N\Delta+\frac{1}{n}(\sqrt{n}a_n)'\widetilde{B}'\widetilde{B}(\sqrt{n}a_n)-\frac{1}{\sqrt{n}}(\Delta X')[\Delta'][\widetilde{W}](\Delta X)+\gamma\|\Delta X\|^2\end{aligned}$$

当 $n\to\infty$ 时，上式中除第一项外都趋于零. 因此(4-1-11)式成立. 由(4-1-11)式可推出(4-1-10)式. 由(4-1-10)式可得到(4-1-9)式.

另外，由于 $S_T\Delta$ 和 $S_N\Delta$ 互相独立，因此由(4-1-11)式和(4-1-12)式可知 S^2 与 $\widetilde{B}(\hat{X}-\widetilde{X})$ 渐近独立. 证毕.

2. 有偏性

以上讨论了非线性最小二乘估计的渐近性质. 那些讨论实际上相当于是在对非线性模型线性近似的基础上进行的. 其结果显然不够精确，甚至还掩盖了某些重要性质. 因此，为了进一步研究非线性最小二乘估计量的统计性质，必须对 $\Delta X=\hat{X}-\widetilde{X}$ 进一步展开，取至二次项. 为书写方便，分别用下列符号表示 $V'V$ 对 X 的一、二、三阶偏导数：

$$\begin{aligned}&\dot{V}=\left.\frac{\partial V'V}{\partial X}\right|_{X=\hat{X}},\ \ddot{V}=\left.\frac{\partial^2 V'V}{\partial X^2}\right|_{X=\hat{X}},\ \dddot{V}=\left.\frac{\partial^3 V'V}{\partial X^3}\right|_{X=\hat{X}}\\&\dot{\widetilde{V}}=\left.\frac{\partial V'V}{\partial X}\right|_{X=\widetilde{X}},\ \ddot{\widetilde{V}}=\left.\frac{\partial^2 V'V}{\partial X^2}\right|_{X=\widetilde{X}},\ \dddot{\widetilde{V}}=\left.\frac{\partial^3 V'V}{\partial X^3}\right|_{X=\widetilde{X}}\end{aligned} \tag{4-1-13}$$

则

$$\dot{V}'=\dot{\widetilde{V}}'+\ddot{\widetilde{V}}\Delta X+\frac{1}{2}\Delta X'\dddot{\widetilde{V}}\Delta X+\gamma\|\Delta X\|^2=0 \tag{4-1-14}$$

因为

$$\dot{\widetilde{V}}'=-2\widetilde{B}'\Delta,\ \ddot{\widetilde{V}}\Delta X=2(\widetilde{B}'\widetilde{B}-[\Delta][W])\Delta X$$

$$\begin{aligned}\frac{1}{2}\Delta X'\dddot{\widetilde{V}}\Delta X&=[(\widetilde{B}\Delta X)'][\widetilde{W}]\Delta X+\frac{1}{2}\Delta X'[\widetilde{B}'][\widetilde{W}]\Delta X\\&\quad-\frac{1}{2}\Delta X'\left\{\Delta'\left(\frac{\partial^3 f}{\partial X^3}\right)\right\}\Delta X\end{aligned}$$

将以上各式代入(4-1-14)式，得

$$\begin{aligned}\widetilde{B}'\widetilde{B}\Delta X&=\widetilde{B}'\Delta+[\Delta'][\widetilde{W}]\Delta X-[(\widetilde{B}X)'][\widetilde{W}]\Delta X-\frac{1}{2}\Delta X'[\widetilde{B}'][\widetilde{W}]\Delta X\\&\quad+\frac{1}{2}\Delta X'\left\{\Delta'\left(\frac{\partial^3 f}{\partial X^3}\right)\right\}\Delta X-\frac{1}{2}\gamma\|\Delta X\|^2\end{aligned}$$

于是有

$$\begin{aligned}\Delta X=\hat{X}-\widetilde{X}&=(\widetilde{B}'\widetilde{B})^{-1}\widetilde{B}'\Delta+(\widetilde{B}'\widetilde{B})^{-1}\{[(\Delta-\widetilde{B}\Delta X)'][\widetilde{W}]\Delta X\\&\quad-\frac{1}{2}\Delta X'[\widetilde{B}'][\widetilde{W}]\Delta X\}+O_t(n^{-3/2})\end{aligned} \tag{4-1-15}$$

(4-1-15)式就是 $\Delta X=\hat{X}-\widetilde{X}$ 的二阶展开式. 当 $n\to\infty$ 时，余项 $O_t(n^{-3/2})$ 趋近于0.

(4-1-15)式右端仍含有 ΔX, 为了消除之, 我们将(4-1-4)式代入(4-1-14)式, 并将与 a_n 有关的项并入 $O_t(n^{-3/2})$ 中, 于是可以得到

$$\Delta X=\hat{X}-\tilde{X}=(\tilde{B}'\tilde{B})^{-1}\tilde{B}'\Delta+(\tilde{B}'\tilde{B})^{-1}\{[(\Delta-\tilde{B}\Delta X)'][\widetilde{W}]\Delta X$$
$$-\frac{1}{2}\Delta'[\tilde{B}'][\tilde{B}(\tilde{B}'\tilde{B})^{-1}\widetilde{W}(\tilde{B}'\tilde{B})^{-1}\tilde{B}']\}+O_t(n^{-3/2}) \qquad (4\text{-}1\text{-}16)$$

(4-1-16)式给出了参数 X 的最小二乘估计 $\hat{X}$ 与随机误差项 Δ 之间的关系. 由于 $L=f(X)+\Delta$, 所以(4-1-16)式同时建立了非线性最小二乘估计 $\hat{X}$ 与观测值 L 之间的关系. 这一关系对分析研究非线性最小二乘估计的性质是很重要的.

定理 4-1-3 若非线性模型(2-1-4)式满足正则条件(ⅰ)~(ⅴ), 则 $\Delta X=\hat{X}-\tilde{X}$ 的二阶展开式可由曲率立体阵 G 和 H 表示为

$$\Delta X=\hat{X}-\tilde{X}=M\tau+M\{[\lambda'][G]\tau-\frac{1}{2}\tau'H\tau\}+O_t(n^{-3/2}) \qquad (4\text{-}1\text{-}17)$$

式中:
$$\tau=Q'\Delta,\quad \lambda=N'\Delta \qquad (4\text{-}1\text{-}18)$$
矩阵 Q,N 同(2-3-1)式. N 的列为 $\tilde{X}$ 处法空间的标准正交基. M,G,H 等都在 $\tilde{X}$ 处计算.

证明: 由于

$$(\tilde{B}'\tilde{B})^{-1}=MM',\quad \tilde{B}'\Delta=R'Q'\Delta=R'\tau$$

因此

$$(\tilde{B}'\tilde{B})^{-1}\tilde{B}'\Delta=MM'\tilde{B}'\Delta=M\Delta$$

式中: M 由(2-3-2)式确定. 将其代入(4-1-16)式, 得

$$\Delta X=\hat{X}-\tilde{X}=M\tau+MM'\{[\Delta'S_N][\widetilde{W}]M\tau-\frac{1}{2}\tau'M'[\tilde{B}'][\widetilde{W}]M\tau+O_t(n^{-3/2})$$

由(2-3-4)式和(2-3-7)式知

$$U=M'\widetilde{W}M,\quad S_N=NN'$$

顾及方括号乘法的运算规则及立体阵的定义可得

$$MM'[\Delta'S_N][\widetilde{W}]M\tau=M[\Delta'NN'][M'\widetilde{W}M]\tau$$
$$=M[\lambda'][[N'][U]]\tau=M[\lambda'][G\tau]$$
$$MM'\left\{-\frac{1}{2}\tau'M'[\tilde{B}'][\widetilde{W}']M\tau\right\}=-\frac{1}{2}MM'\{\tau'[\tilde{B}'][U]\tau\}$$
$$=-\frac{1}{2}M\{\tau'[M'\tilde{B}'][U]\tau\}$$
$$=-\frac{1}{2}M\{\tau'[Q'][U]\tau\}=-\frac{1}{2}M(\tau'H\tau)$$

将以上各式代入(4-1-16)式, 即得(4-1-17)式.

定理 4-1-4 若非线性模型(2-1-4)式满足正则条件(ⅰ)~(ⅴ), 则非线性最小二乘估计为有偏估计, 其偏差为

$$\mathrm{Bias}(\hat{X})=-\frac{\sigma^2}{2}M\mathrm{tr}(H)+O(n^{-2}) \qquad (4\text{-}1\text{-}19)$$

证明: 对(4-1-17)式取消数学期望, 并顾及(4-1-18)式得

$$\mathrm{E}(\hat{X}-\tilde{X})=MQ'\mathrm{E}(\Delta)+M\mathrm{E}([\lambda'][G]\tau)-\frac{1}{2}M\mathrm{E}(\tau'H\tau)$$

因为 $\mathrm{E}(\Delta)=0$

$$E([\lambda'][G]\tau)=\sum_{i=1}^{n-t}E(\lambda_i G_i \tau)=\sum_{i=1}^{n-t}E(\lambda_i)G_i E(\tau)=0,$$

而
$$E(\tau' H_i \tau)=E(\mathrm{tr}(H_i \tau\tau'))=\sigma^2 \mathrm{tr}(H_i)$$

式中 H_i 为 H 的第 i 层，又 $\mathrm{tr}(H)=(\mathrm{tr}(H_1),\mathrm{tr}(H_2),\cdots,\mathrm{tr}(H_i))'$，所以

$$E(\hat{X}-\tilde{X})=-\frac{\sigma^2}{2}M\mathrm{tr}(H)+O(n^{-2})$$

(4-1-19)式表明，非线性最小二乘估计 $\hat{X}$ 的偏差与非线性模型的固有曲率无关，仅与参数效应曲率有关.这就是说非线性最小二乘估计 $\hat{X}$ 的偏差取决于参数的选择.另外，由于模型的固有非线性强度通常较弱，因此，非线性最小二乘估计 $\hat{X}$ 的偏差 $\mathrm{Bias}(\hat{X})$ 可以用来衡量非线性模型的非线性强度.一般用 $\hat{X}$ 的各分量的百分偏差衡量之，$\hat{X}$ 的各分量的百分偏差记为

$$\%\,\mathrm{Bias}(\hat{X}_i)=100\,\frac{\mathrm{Bias}(\hat{X}_i)}{\hat{X}_i}\,\% \tag{4-1-20}$$

为了用(4-1-20)式衡量非线性模型的非线性强度，Ratkowsky 曾给了一个经验法则，认为当百分偏差超过1%，就说明非线性模型强度较强.

由于 σ^2 和 $\tilde{X}$ 都未知，故按(4-1-19)式无法计算偏差.因此，实用上总是用 $\hat{\sigma}^2$ 和 $\hat{X}$ 代替 σ^2 和 $\tilde{X}$，故偏差的实用形式为

$$\mathrm{Bias}(\hat{X})=-\frac{\hat{\sigma}^2}{2}(B'B)^{-1}B'\mathrm{tr}((B'B)^{-1}W) \tag{4-1-21}$$

证明：略去高阶项，并由立体阵的性质知

$$\mathrm{Bias}(\hat{X})\approx-\frac{\sigma^2}{2}M\mathrm{tr}(H)=-\frac{\sigma^2}{2}\mathrm{tr}([M][H])=-\frac{\sigma^2}{2}\mathrm{tr}([M][Q'][U])$$
$$=-\frac{\sigma^2}{2}\mathrm{tr}([MQ'][U])=-\frac{\sigma^2}{2}MQ'\mathrm{tr}(M'\widetilde{W}M)$$
$$=-\frac{\sigma^2}{2}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'\mathrm{tr}(MM'\widetilde{W})$$
$$=-\frac{\sigma^2}{2}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'\mathrm{tr}((\tilde{B}'\tilde{B})^{-1}\widetilde{W})$$

用估值代替真值得

$$\mathrm{Bias}(\hat{X})\approx-\frac{\hat{\sigma}^2}{2}(B'B)^{-1}B'\mathrm{tr}(B'B)^{-1}W$$

以上推导用到了 $MM'=(B'B)^{-1}$

$$MQ'=MM'R'Q'=(B'B)^{-1}B'$$

3.方差非最小性

为了讨论这一性质，我们先证明两个引理.

引理 4-1-1 由(4-1-18)式定义的 τ 和 λ 均为正态变量，$\tau\sim N(0,\sigma^2 I)$，$\lambda\sim N(0,\sigma^2 I)$，且 τ 与 λ 相互独立.于是下列各式成立.

$$E(\tau_i\tau_j\tau_k)=0,\quad E(\tau_i\tau_j\tau_k\tau_l\tau_m)=0 \tag{4-1-22}$$
$$E(\tau_i\tau_j\tau_k\tau_l)=\sigma^4(\delta_{ij}\delta_{kl}+\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}) \tag{4-1-23}$$
$$E(\lambda_i\tau_j\tau_k)=0,\quad E(\lambda_i\tau_j\tau_k\tau_l)=0 \tag{4-1-24}$$

式中：$\delta_{ij}=1(i=j)$，$\delta_{ij}=0(i\neq j)$

证明：由于(Q,N)为正交矩阵，因此根据 τ 与λ 的定义有$\begin{pmatrix}\tau\\ \lambda\end{pmatrix}=\begin{pmatrix}Q'\\ N'\end{pmatrix}\Delta\sim N(0,\sigma^2I)$. 即$(\tau',\lambda')$与 Δ 同分布. 因此，τ 和λ 皆为正态变量且相互独立. 显然(4-1-22)式和(4-1-24)式成立. 而(4-1-23)式是标准正态变量的四阶矩公式，所以引理得证.

引理 4-1-2 设 S、T 分别为$n\times t\times t$ 阶和 $m\times t\times t$ 阶立体阵. 它们的每层都是对称矩阵，又设 τ 为t 阶标准正态随机向量，即 $\tau\sim N(0,\sigma^2I)$，则有

$$E(\tau'S\tau)=\sigma^2\mathrm{tr}(S)$$

$$\mathrm{Cov}(\tau'S\tau,\tau'T\tau)=2\sigma^4\sum_{k=1}^{t}\sum_{l=1}^{t}S_{kl}T'_{kl}=2\sigma^4(\mathrm{tr}(S_iT_j))$$

式中：S_{kl}和 T_{kl}为立体阵S、T 在(k,l)处的 n 维及 m 维向量. $(\mathrm{tr}(S_iT_j))$表示在(i,j)处元素为 $\mathrm{tr}(S_iT_j)$的 $n\times m$ 阶矩阵.

证明：$E(\tau'S\tau)$的第 i 个元素为$E(\tau'S_i\tau)=\sigma^2\mathrm{tr}(S_i)$，故由 $\mathrm{tr}(S)$的定义知 $E(\tau'S_i\tau)=\sigma^2\mathrm{tr}(S)$成立.

又因为 $\mathrm{Cov}(\tau'S\tau,\tau'T\tau)$在$(i,j)$处的元素为

$$\mathrm{Cov}(\tau'S\tau,\tau'T\tau)=E\left(\left(\sum_{k=1}^{t}\sum_{l=1}^{t}S_{ikl}\tau_k\tau_l\right)\left(\sum_{s=1}^{t}\sum_{e=1}^{t}T_{jse}\tau_s\tau_e\right)\right)-(E(\tau'S\tau))(\tau'T_j\tau)$$

$$=\sum_{k,l}\sum_{s,e}S_{ike}T_{jse}(\delta_{ki}\delta_{se}+\delta_{ks}\delta_{le}+\delta_{ke}\delta_{se})\sigma^4-\sigma^4(\mathrm{tr}S_i)(\mathrm{tr}Q_j)$$

$$=2\sigma^4\sum_{k-1}^{t}\sum_{l-1}^{t}S_{ike}T_{jke}$$

于是有 $\mathrm{Cov}(\tau'S\tau,\tau'T\tau)=2\sigma^4\sum_{k=e}^{t}\sum_{l=1}^{t}S_{kl}T'_{kl}$

定理 4-1-5 设非线性模型(2-1-4)式满足正则条件(ⅰ)~(ⅴ)，则非线性最小二乘估计 $\hat{X}$ 的方差达不到最小值(CR 下界)，且可由曲率立体阵表示为

$$\mathrm{Var}(\hat{X})\approx\sigma^2(B'B)^{-1}+\frac{1}{2}\sigma^4M(2A_G+A_H)M' \tag{4-1-25}$$

式中：

$$A_G=\sum_{s=1}^{n-t}G_s^2 \tag{4-1 26}$$

$$A_H=\sum_{s=1}^{t}\sum_{l=1}^{t}H_{sl}H'_{sl}=(\mathrm{tr}(H_iH_j)) \tag{4-1-27}$$

M 的定义同前. G_s 为固有曲率立体阵G 的第s 层；H_i 表示参数效应曲率立体阵H 的第i层；H_{se}表示H 在(s,e)处的 t 维向量；$A_H=\mathrm{tr}(H_iH_j)$表示 A_H 的第i 行第j 列元素为 $\mathrm{tr}(H_iH_j)$.

证明：略去(4-1-17)式中的高阶余项后，应用协方差传播律，得

$$\mathrm{Var}(\hat{X})=M\{\mathrm{Var}(\tau)+\mathrm{Var}([\lambda'][G]\tau)+\frac{1}{4}\mathrm{Var}(\tau'H\tau)+\text{交叉项}\}M' \tag{4-1-28}$$

由引理 4-1-1 知，各交叉项皆为零. 例如 $\mathrm{Cov}(\tau,[\lambda'][G]\tau)$在$(i,j)$处的元素为

$$E\left(\tau\sum_{j=1}^{n-t}\sum_{k=1}^{t}\lambda_jG_{jlk}\tau_k\right)=\sum_{j=1}^{n-t}\sum_{k=1}^{t}G_{jek}E(\tau_i\lambda_j\tau_k)=0$$

由引理知

$$\mathrm{Cov}([\lambda'][G]\tau,\tau'H\tau)=0$$
$$\mathrm{Cov}(\tau,\tau'H\tau)=0$$

再看(4-1-28)式的前三项

$$M\mathrm{Var}(\tau)M'=\sigma^2MM'=\sigma^2(B'B)^{-1}$$

这就是(4-1-25)式的第一项.而

$$\begin{aligned}\mathrm{Var}([\lambda'][G]\tau)&=\mathrm{Var}(\sum_{s=1}^{n-t}\lambda_sG_s\tau)\\&=\sum_{s=1}^{n-t}\mathrm{Var}(\lambda_sG_s\tau)+\text{交叉项}\end{aligned}$$

因为交叉项为零,故由 τ 与λ_k 的独立性可得

$$\begin{aligned}\mathrm{Var}([\lambda'][G]\tau)&=\sum_{s=1}^{n-t}\sigma^2\mathrm{Var}(G_s\tau)=\sigma^2\sum_{s=1}^{n-t}G_s\mathrm{Var}(\tau)G_s'\\&=\sigma^4\sum_{s=1}^{n-t}G_s^2=\sigma^4A_G\end{aligned}$$

于是 $M\mathrm{Var}([\lambda'][G]\tau)M'=\sigma^4MA_GM'$.这就是(4-1-25)式的第二项.

对于最后一项,因 $\mathrm{Var}(\tau'H\tau)$在(i,j)处的元素为 $\mathrm{Cov}(\tau'H_i\tau,\tau'H_j\tau)$.由 τ 的正态性以及H_i 和H_j 的对称性,并顾及引理 4-1-2,得

$$\mathrm{Cov}(\tau'H_i\tau,\tau'H_i\tau)=2\sigma^4\mathrm{tr}(H_iH_j)=2\sigma^4\sum_{k=1}^{t}\sum_{l=1}^{t}H_{ikl}H_{jkl}$$

故$\frac{1}{4}\mathrm{Var}(\tau'H\tau)=\frac{1}{2}\sigma^4\sum_{k=1}^{t}\sum_{l=1}^{t}H_{kl}H_{kl}'=\frac{1}{2}\sigma^4A_H$.于是$\frac{1}{4}M\mathrm{Var}(\tau'H\tau)M'=\frac{1}{2}\sigma^4MA_HM'$,这就是(4-1-25)式的第三项.

由于观测向量 L 的分布关于X 的 Fisher 信息阵为 $\sigma^{-2}(B'B)$.所以(4-1-25)式的第一项即为 CR 下界,而 $MA_GM'=\sum_{s=1}^{n-t}MG_s^2M'$, $MA_HM'=\sum_{k=1}^{t}\sum_{l=1}^{t}MH_{kl}'H_{kl}'M'$都是非负矩阵.故只要立体阵 G 和H 不恒为零,则(4-1-25)式的第二、第三项中的两个矩阵都是正定矩阵.于是(4-1-25)式中的第二、第三两项都大于 0.因此,非线性模型参数的最小二乘估计 $\hat{X}$ 的方差一定大于CR 下界(即 $\mathrm{Var}(\hat{X})>\sigma^2(B'B)^{-1}$).所以非线性最小二乘估计 $\hat{X}$ 的方差达不到最小值.定理证毕.

第二节　单位权方差 σ^2 的估计

因为残差向量的平方和 $V'V$ 中包含了单位权方差σ^2 的信息,所以为了得到单位权方差 σ^2 的估值 $\hat{\sigma}^2$,有必要先讨论非线性最小二乘估计中残差向量 V 的展开式及其期望和方差[43](王新洲,2000).

一、残差向量 V 的展开式

定理 4-2-1　设非线性模型(2-1-4)式满足正则条件(ⅰ)~(ⅴ),则残差向量 V 可展开为

$$V=\Delta f-\Delta=-N\lambda+Q[\lambda'][G]\tau+\frac{1}{2}N(\tau'G\tau)+O_t(n^{-1}) \tag{4-2-1}$$

式中：
$$\Delta f=f(\hat{X})-f(\tilde{X})=Q\tau+[\lambda'][G]\tau+\frac{1}{2}N(\tau'G\tau)+O_t(n^{-1}) \tag{4-2-2}$$

称为拟合误差向量.

(4-2-1)式与(4-2-2)式仅与非线性模型的固有曲率有关,在参数变换下保持不变.

证明：先证明(4-2-2)式.将 Δf 在 $\hat{X}$ 处展开可得

$$\Delta f=\frac{\partial f}{\partial X}\Delta X+\frac{1}{2}\Delta X'\frac{\partial^2 f}{\partial X^2}\Delta X+O_t(\|\Delta X\|^3)$$

$$=B\Delta X+\frac{1}{2}\Delta X'W\Delta X+O_t(\|\Delta X\|^3) \tag{4-2-3}$$

将(4-1-17)式两边左乘 B,得(4-2-3)式的第一项

$$B\Delta X=BM\tau+BM\{[\lambda'][G]\tau-\frac{1}{2}\tau'H\tau\}+O_t(n^{-1})$$

$$=Q\tau+Q\{[\tau'][G]\tau-\frac{1}{2}\tau'H\tau\}+O_t(n^{-1}) \tag{4-2-4}$$

又因为 $\Delta X=\overline{\Delta X}+\alpha$, $\overline{\Delta X}=(B'B)^{-1}B'\Delta=MQ'\Delta=M\tau=O_t(n^{-1/2})$, $\alpha=O_t(n^{-1})$,于是(4-2-3)式的第二项可写为

$$\frac{1}{2}\Delta X'W\Delta X=\frac{1}{2}\overline{\Delta X}'W\overline{\Delta X}+\frac{1}{2}\alpha'W\alpha+\overline{\Delta X}'W\alpha$$

$$=\frac{1}{2}\tau'M'WM\tau+\frac{1}{2}n^{-2}(n\alpha)'W(n\alpha)+n^{-3/2}(\sqrt{n\Delta X})'W(n\alpha)$$

$$=\frac{1}{2}\tau'U\tau+O_t(n^{-1}) \tag{4-2-5}$$

将(4-2-4)式和(4-2-5)式代入(4-2-3)式,得

$$\Delta f-Q\tau+Q[\lambda'][G]\tau+\frac{1}{2}\tau'U\tau-\frac{1}{2}Q(\tau'H\tau)+O_t(n^{-1}) \tag{4-2-6}$$

由立体阵的运算性质,并顾及 $U=[N][G]+[Q][H]$,则(4-2-6)式的第三、四项可合并为

$$\frac{1}{2}\tau'U\tau-\frac{1}{2}Q(\tau'H\tau)=\frac{1}{2}\tau'U\tau-\frac{1}{2}\tau'[Q][H]\tau$$

$$=\frac{1}{2}\tau'\{U-[Q][H]\}\tau=\frac{1}{2}\tau'[N][G]\tau=\frac{1}{2}N(\tau'G\tau)$$

将其代入(4-2-6)式,即得

$$\Delta f=Q\tau+Q[\lambda'][G]\tau+\frac{1}{2}N(\tau'G\tau)+O_t(n^{-1})$$

下面再来证明(4-2-1)式.

因为

$$V=Q\tau-\Delta+Q[\tau'][G]\tau+\frac{1}{2}N(\tau'G\tau)+O_t(n^{-1})$$

而前两项为

$$Q\tau-\Delta=QQ'\Delta-\Delta=-(I-QQ')\Delta=-S_N\Delta=-NN'\Delta=-N\lambda$$

所以有

$$V = -N\lambda + Q[\lambda'][G]\tau + \frac{1}{2}N(\tau'G\tau) + O_t(n^{-1}).$$

由于(4-2-1)式和(4-2-2)式中仅出现固有曲率立体阵 G,而与参数效应曲率无关,所以 V 和 Δf 的展开式在参数变换下保持不变.定理证毕.

二、残差向量 V 的期望和方差

对(4-2-1)式两边取期望得

$$E(V) = -NE(\lambda) + QE([\lambda'][G]\tau) + \frac{1}{2}NE(\tau'G\tau) + O_t(n^{-1})$$

因为 $E(\lambda) = E(N'\Delta) = N'E(\Delta) = 0$. 又由定理 4-1-4 和引理 4-1-2 的证明知

$$E([\lambda'][G]\tau) = 0, E(\tau'G\tau) = \sigma^2 tr(G)$$

所以残差向量 V 的数学期望为

$$E(V) = \frac{\sigma^2}{2}N tr(G) + O_t(n^{-1}) \tag{4-2-7}$$

再对(4-2-1)式两端取方差,得

$$Var(V) = Var(N\lambda) + Var(Q[\lambda'][G]\tau) + \frac{1}{4}Var(N(\tau'G\tau)) + \text{交叉项} \tag{4-2-8}$$

由定理 4-1-5 的证明知交叉项皆为零.再由协方差传播律知

$$\begin{aligned} Var(N\lambda) &= NVar(\lambda)N' = NVar(N'\Delta)N' \\ &= \sigma^2 NN'NN' = \sigma^2 NN' = \sigma^2 S_N \end{aligned} \tag{4-2-9}$$

$$Var(Q[\lambda'][G]\tau) = QVar([\lambda'][G]\tau)Q' = \sigma^4 QA_GQ' \tag{4-2-10}$$

$$\begin{aligned} \frac{1}{4}Var(N(\tau'G\tau)) &= \frac{1}{4}NVar(\tau'G\tau)N' = \frac{1}{2}\sigma^4 N(\sum_{k=1}^{t}\sum_{l=1}^{t} G_{kl}G'_{kl})N' \\ &= \frac{1}{2}\sigma^4 NA_G^*N' \end{aligned} \tag{4-2-11}$$

式中:

$$A_G^* = \sum_{k=1}^{t}\sum_{l=1}^{t} G_{kl}G'_{kl} \tag{4-2-12}$$

以上推导用到了

$$\left.\begin{aligned} Var([\lambda'][G]\tau) &= \sigma^4 A_G \\ Var(\tau'G\tau) &= 2\sigma^4 A_G^* \end{aligned}\right\} \tag{4-2-13}$$

(4-2-13)式的推导详见定理 4-1-5 的证明.

将(4-2-9)式、(4-2-10)式和(4-2-11)式代入(4-2-8)式,得残差向量的方差为

$$Var(V) = \sigma^2 S_N + \sigma^4\left(QA_GQ' + \frac{1}{2}NA_A^*N'\right) \tag{4-2-14}$$

三、单位权方差 σ^2 的估计

和线性模型中的情况一样,我们仍然用残差平方和 $V'V$ 来估计单位权方差 σ^2. 为此,对残差平方和 $V'V$ 取期望.由(1-1-21)式知,二次型 $V'V$ 的期望为

$$E(V'V) = (E(V))'(E(V)) + tr(Var(V)) \tag{4-2-15}$$

由(4-2-7)式知

$$(\mathrm{E}(V))'(\mathrm{E}(V))=\left(\frac{\sigma^2}{2}N\mathrm{tr}(G)\right)'\left(\frac{\sigma^2}{2}N\mathrm{tr}(G)\right)=\frac{\sigma^4}{4}(\mathrm{tr}(G))'N'N\mathrm{tr}(G)$$
$$=\frac{\sigma^4}{4}(\mathrm{tr}(G))'\mathrm{tr}(G)=\frac{\sigma^4}{4}\|\mathrm{tr}(G)\|^2 \tag{4-2-16}$$

再由(4-2-14)式知

$$\mathrm{tr}(\mathrm{Var}(V))=\sigma^2\mathrm{tr}(S_N)+\sigma^2\mathrm{tr}\left(QA_GQ'+\frac{1}{2}NA_G^*N'\right)$$
$$=\sigma^2\mathrm{tr}(I-B(B'B)^{-1}B')+\sigma^4\mathrm{tr}\left(A_GQ'Q+\frac{1}{2}A_G^*N'N\right)$$
$$=\sigma^2(n-t)+\sigma^4\left(\mathrm{tr}(A_G)+\frac{1}{2}\mathrm{tr}(A_G^*)\right)$$

因为

$$\mathrm{tr}(A_G)=\sum_{e=1}^{n-t}\mathrm{tr}(G_e^2)=\sum_{e=1}^{n-t}\sum_{k=1}^{t}\sum_{l=1}^{t}G_{kel}^2=\sum_{k=1}^{t}\sum_{l=1}^{t}\|G_{kl}\|^2$$
$$\mathrm{tr}(A_G^*)=\sum_{k=1}^{t}\sum_{e=1}^{t}\mathrm{tr}(G_{ke}'G_{ke})=\sum_{k=1}^{t}\sum_{e=1}^{t}\|G_{ke}\|^2$$

于是有

$$\mathrm{tr}(\mathrm{Var}(V))=\sigma^2(n-t)+\frac{3}{2}\sigma^4\sum_{k=1}^{t}\sum_{e=1}^{t}\|G_{ke}\|^2 \tag{4-2-17}$$

将(4-2-16)式和(4-2-17)式同时代入(4-2-15)式，得

$$\mathrm{E}(V'V)=\sigma^2(n-t)+\sigma^4\left(\frac{3}{2}\sum_{k=1}^{t}\sum_{e=1}^{t}\|G_{ke}\|^2+\frac{1}{4}\|\mathrm{tr}(G)\|^2\right) \tag{4-2-18}$$

去掉(4-2-18)式中的期望，用估值 $\hat{\sigma}^2$ 代替真值 σ^2，顾及 $r=n-t$，并令

$$a=\frac{3}{2}\sum_{k=1}^{t}\sum_{e=1}^{t}\|G_{ke}\|^2+\frac{1}{4}\|\mathrm{tr}(G)\|^2 \tag{4-2-19}$$

可得

$$a\hat{\sigma}^4+r\hat{\sigma}^2-V'V=0 \tag{4-2-20}$$

当模型(2-1-4)式为线性模型时，有 $G=0$，于是 $a=0$，则(4-2-20)式变为 $r\hat{\sigma}^2=V'V$. 故有

$$\hat{\sigma}^2=\frac{V'V}{r}=S^2$$

此式正是线性模型中单位权方差的估计公式. 这说明(4-2-20)式与线性模型的结果完全一致. 这也正是我们所期望的.

对于非线性模型(2-1-4)式，有 $G\neq0$，故 $a\neq0$. 于是(4-2-20)式就是关于 $\hat{\sigma}^2$ 的一元二次方程. 应用解一元二次方程的公式法解(4-2-20)式，得

$$\hat{\sigma}^2=\frac{-r\pm\sqrt{r^2+4aV'V}}{2a} \tag{4-2-21}$$

(4-2-21)式有两个根 $\hat{\sigma}^2=\frac{-r+\sqrt{r^2+4aV'V}}{2a}$ 和 $\hat{\sigma}^2=\frac{-r-\sqrt{r^2+4aV'V}}{2a}$. 因为 $a>0$, $r>0$, $V'V>0$，故 $\hat{\sigma}^2=\frac{-r-\sqrt{r^2+4aV'V}}{2a}<0$. 这显然不符合实际，应舍去. 所以非线性模型参数估计中单位权方差 σ^2 的严密估计公式为[43]（王新洲，2000）:

$$\hat{\sigma}^2=\frac{-r+\sqrt{r^2+4aV'V}}{2a} \tag{4-2-22}$$

由(4-2-19)式知，a 仅与非线性模型的固有曲率立体阵 G 有关，而与参数效应曲率无关，所以单位权方差的估值 $\hat{\sigma}^2$ 在参数变换下保持不变. 这是必然的，因为参数变换并不改变观测值的精度.

例 4-2-1 设非线性模型为 $L_i=x^{i+1}+\Delta_i$，参数的真值为 $\tilde{x}=2.013564$，模拟的观测值 L_i，观测值的真值 $\tilde{L}_i$ 和真误差 Δ_i 列于表 4-2-1. 试用(4-2-22)式估计单位权方差 σ^2.

表 4-2-1

i	$\tilde{L}_i$	L_i	Δ_i
1	4.054439982	4.04	−0.014439982
2	8.163874388	8.18	0.016125612
3	16.438483570	16.45	0.011516433

取参数的近似值为 $x_0=2$，用高斯-牛顿法经 4 次迭代，得

$$B=\begin{bmatrix}4.027\ 963\\12.168\ 366\\32.675\ 821\end{bmatrix},\quad V=\begin{bmatrix}0.016\ 121\ 999\\-0.011\ 044\ 810\\0.002\ 125\ 670\end{bmatrix},\quad W=\begin{bmatrix}2\\12.083890\\48.673464\end{bmatrix}$$

$$\hat{x}=2.013981629,\quad V'V=0.000386425$$

$$R=\sqrt{B'B}=35.099898,\quad M=R^{-1}=0.028490$$

$$Q=BM=\begin{bmatrix}0.114\ 757\\0.346\ 678\\0.930\ 938\end{bmatrix},\quad N=\begin{bmatrix}0.115 & -0.986714\\-0.935464 & -0.068717\\0.334118 & 0.147222\end{bmatrix}$$

$$G=\begin{bmatrix}0.004\ 210\\0.003\ 541\end{bmatrix},a=\left(\frac{3}{2}+\frac{1}{4}\right)(0.004210^2+0.003541^2)=0.000053$$

将 a，r 和 $V'V$ 代入(4-2-22)式，得

$$\hat{\sigma}^2=0.000192630,\hat{\sigma}=\pm0.013879$$

而根据 $\sigma^2=\frac{\Delta'\Delta}{n}$ 得

$$\sigma^2=0.000200392,\sigma=\pm0.014156$$

按线性近似得

$$S^2=\frac{V'V}{r}=0.000193213,S=\pm0.013900$$

此例的固有曲率立体阵 G 的数值较小，非线性强度较弱. 下面再看一个固有曲率立体阵 G 的数值较大的例子.

例 4-2-2 非线性模型为 $L_i=e^{ix}+\Delta_i$. 参数的真值 $\tilde{x}=-0.25413679$. 模拟观测值，观测值的真值和真误差列于表 4-2-2. 试用(4-2-22)式估计单位权中误差 σ^2.

表 4-2-2

i	$\hat{L}_i$	L_i	Δ_i
1	0.775585702	0.79	0.01441429
2	0.601533181	0.61	0.008466818
3	0.466540535	0.45	−0.01654053

取 $x_0 = -0.255$，用高斯-牛顿法迭代 5 次，得

$$B = \begin{bmatrix} 0.775\ 243 \\ 1.202\ 004 \\ 1.397\ 768 \end{bmatrix}, \quad V = \begin{bmatrix} 0.016\ 121\ 999 \\ -0.011\ 044\ 810 \\ 0.002\ 125\ 670 \end{bmatrix}, \quad W = \begin{bmatrix} 2 \\ 12.083890 \\ 48.673464 \end{bmatrix}$$

$$\hat{x} = -0.25457862, \quad V'V = 0.000552274$$

$$R = \sqrt{B'B} = 1.999892, \quad M = 0.500027$$

$$Q = BM = \begin{bmatrix} 0.387\ 642 \\ 0.601\ 034 \\ 0.698\ 921 \end{bmatrix}, \quad N = \begin{bmatrix} 0.1 & -0.961371 \\ -0.781150 & 0.169006 \\ 0.616283 & 0.362911 \end{bmatrix}$$

$$G = \begin{pmatrix} 0.195\ 995 \\ 0.304\ 452 \end{pmatrix}, \quad a = 0.229434$$

将 a，r 和 $V'V$ 代入(4-2-22)式，得

$$\hat{\sigma}^2 = 0.000276128, \quad \hat{\sigma} = \pm 0.016617106$$

而

$$\sigma^2 = \frac{\Delta'\Delta}{n} = 0.000184349, \quad \sigma = \pm 0.013577534$$

$$S^2 = \frac{V'V}{r} = 0.000276137, \quad S = \pm 0.01661737$$

由上两例可以看出 $S^2 > \hat{\sigma}^2$，这是因为由 S^2 的定义知

$$rS^2 - V'V = 0 \tag{4-2-23}$$

将(4-2-23)式减去(4-2-20)式，得

$$r(S^2 - \hat{\sigma}^2) = a\hat{\sigma}^4$$

因为 $a > 0$，所以

$$r(S^2 - \hat{\sigma}^2) = a\hat{\sigma}^4 > 0$$

即

$$S^2 > \hat{\sigma}^2 \tag{4-2-24}$$

(4-2-24)式表明，在非线性模型参数估计中，按近似公式(4-2-23)式估计单位权方差 σ^2 比按严密公式(4-2-22)式估计的结果大．但 $\hat{\sigma}^2$ 与 S^2 相差很小很小．而按(4-2-22)式计算 $\hat{\sigma}^2$，其工作量非常之大．加之实际测量中随机误差总是微小量，单位权中误差取至百分位精度足够了，何况用 S^2 比用 $\hat{\sigma}^2$ 更安全．所以建议在实际工作中仍用 $S^2 = \frac{V'V}{r}$ 去估计单位权方差 σ^2[43]（王新洲，2000）．

第三节　非线性函数的误差传播

一、观测值函数的方差

设独立观测向量的真值为 $\tilde{L}=(\tilde{l}_1\quad \tilde{l}_2\quad \cdots\quad \tilde{l}_n)'$，观测值为 $L=(l_1\quad l_2\quad \cdots\quad l_n)'$，真误差为 $\Delta=(\Delta_1\quad \Delta_2\quad \cdots\quad \Delta_n)'$服从正态分布，即 $\Delta\sim N(0,\mathrm{Var}(\Delta))$. 其中

$$\mathrm{Var}(\Delta)=\mathrm{Var}(L)=\begin{bmatrix}\sigma_1^2 & 0 & \cdots & 0\\ 0 & \sigma_2^2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \sigma_n^2\end{bmatrix}$$

现有独立观测向量的非线性函数

$$y=\varphi(\tilde{L})+\varphi_0 \tag{4-3-1}$$

式中 φ_0 为常数. 要求根据独立观测向量 L 的方差Var(L)，来求非线性函数 y 的方差 Var(y). 这就是非线性函数的误差传播问题. 为了根据观测向量 L 的方差 Var(L)求非线性函数 y 的方差 Var(y)，将(4-3-1)式在观测值 L 处展为泰勒级数，并取至二次项得：

$$\begin{aligned}y=&y_0+a_1\Delta_1+\cdots+a_n\Delta_n+\frac{1}{2}c_{11}\Delta_1^2+\cdots+\frac{1}{2}c_{nn}\Delta_n^2+c_{12}\Delta_1\Delta_2\\&+\cdots+c_{1n}\Delta_1\Delta_n+c_{23}\Delta_2\Delta_3+\cdots+c_{n-1,n}\Delta_{n-1}\Delta_n\end{aligned}$$

式中：

$$y_0=\varphi(L)+\varphi_0$$

$$a_i=\left.\frac{\partial\varphi}{\partial\tilde{L}_i}\right|_{\tilde{L}=L},\quad c_{ii}=\left.\frac{\partial^2\varphi}{\partial\tilde{L}_i^2}\right|_{\tilde{L}=L},\quad c_{ij}=\left.\frac{\partial^2\varphi}{\partial\tilde{L}_i\partial\tilde{L}_j}\right|_{\tilde{L}=L} \tag{4-3-2}$$

令

$$\begin{aligned}\Delta y=&a_1\Delta_1+\cdots+a_n\Delta_n+\frac{1}{2}c_{11}\Delta_1^2+\cdots+\frac{1}{2}c_{nn}\Delta_n^2+c_{12}\Delta_1\Delta_2\\&+\cdots+c_{1n}\Delta_1\Delta_n+c_{23}\Delta_2\Delta_3+\cdots+c_{n-1,n}\Delta_{n-1}\Delta_n\end{aligned} \tag{4-3-3}$$

对 Δy 取数学期望，并顾及

$$\begin{gathered}\mathrm{E}(\Delta_i)=0,\quad \mathrm{E}(\Delta_i\Delta_j)=0(i\neq j),\quad \mathrm{E}(\Delta_i^3)=0,\quad \mathrm{E}(\Delta_i^2)=\sigma_i^2\\ \mathrm{E}(\Delta_i\Delta_j\Delta_k)=0,\quad \mathrm{E}(\Delta_i\Delta_j^2)=0,\quad \mathrm{E}(\Delta_i^4)=3\sigma_i^4,\quad \mathrm{E}(\Delta_i^2\Delta_j^2)=\sigma_i^2\sigma_j^2\end{gathered} \tag{4-3-4}$$

得

$$\mathrm{E}(\Delta y)=\frac{1}{2}(c_{11}\sigma_1^2+c_{22}\sigma_2^2+\cdots+c_{nn}\sigma_n^2)$$

根据方差的定义知：

$$\begin{aligned}\mathrm{Var}(y)=\sigma_{yn}^2&=\mathrm{E}[(\Delta y-\mathrm{E}(\Delta y))^2]\\&=\mathrm{E}\left[\left(\begin{aligned}&a_1\Delta_1+\cdots+a_n\Delta_n+\frac{1}{2}c_{11}(\Delta_1^2-\sigma_1^2)+\cdots+\frac{1}{2}c_{nn}(\Delta_n^2-\sigma_n^2)+c_{12}\Delta_1\Delta_2\\&+\cdots+c_{n-1,n}\Delta_{n-1}\Delta_n\end{aligned}\right)^2\right]\end{aligned}$$

顾及(4-3-4)式，得

$$Var(y)=\sum_{i=1}^{n}a_i^2\sigma_i^2+\frac{1}{2}\sum_{i=1}^{n}c_{ii}^2\sigma_i^4+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}c_{ij}^2\sigma_i^2\sigma_j^2 \tag{4-3-5}$$

例 4-3-1 在 GIS 矢量数据库中得 A、B 两点在空间直角坐标系中的坐标观测值为 $L=(l_1\quad l_2\quad l_3\quad l_4\quad l_5\quad l_6)'=(x_A\quad y_A\quad z_A\quad x_B\quad y_B\quad z_B)'=(3\quad 5\quad 4\quad 2\quad 3\quad 2)'$. L 的方差阵为:

$$Var(X)=\begin{pmatrix}\sigma_{x_A}^2&0&0&0&0&0\\0&\sigma_{y_A}^2&0&0&0&0\\0&0&\sigma_{z_A}^2&0&0&0\\0&0&0&\sigma_{x_B}^2&0&0\\0&0&0&0&\sigma_{y_B}^2&0\\0&0&0&0&0&\sigma_{z_B}^2\end{pmatrix}=\begin{pmatrix}1/2&0&0&0&0&0\\0&1/2&0&0&0&0\\0&0&1&0&0&0\\0&0&0&1/4&0&0\\0&0&0&0&1/4&0\\0&0&0&0&0&1/2\end{pmatrix}$$

试求空间直线

$$y=\sqrt{(x_B-x_A)^2+(y_B-y_A)^2+(z_B-z_A)^2} \tag{4-3-6}$$

的方差 σ_{ym}^2.

解:将观测值代入(4-3-6)式可算出该空间直线近似值为

$$y_0=\sqrt{(2-3)^2+(3-5)^2+(2-4)^2}=3$$

于是有

$$a_1=\frac{\partial y}{\partial x_1}=\frac{\partial y}{\partial x_A}=\frac{1}{3},\quad a_2=\frac{\partial y}{\partial x_2}=\frac{\partial y}{\partial y_A}=\frac{2}{3},$$

$$a_3=\frac{\partial y}{\partial x_3}=\frac{\partial y}{\partial z_A}=\frac{2}{3},\quad a_4=\frac{\partial y}{\partial x_4}=\frac{\partial y}{\partial x_B}=-\frac{1}{3},$$

$$a_5=\frac{\partial y}{\partial x_5}=\frac{\partial y}{\partial y_B}=-\frac{2}{3},\quad a_6=\frac{\partial y}{\partial x_6}=\frac{\partial y}{\partial z_B}=-\frac{2}{3},$$

$$c_{11}=\frac{\partial^2 y}{\partial x_1^2}=\frac{\partial^2 y}{\partial x_A^2}=\frac{8}{27},\quad c_{22}=\frac{\partial^2 y}{\partial x_2^2}=\frac{\partial^2 y}{\partial y_A^2}=\frac{5}{27},$$

$$c_{33}=\frac{\partial^2 y}{\partial x_3^2}=\frac{\partial^2 y}{\partial z_A^2}=\frac{5}{27},\quad c_{44}=\frac{\partial^2 y}{\partial x_4^2}=\frac{\partial^2 y}{\partial x_B^2}=\frac{8}{27},$$

$$c_{55}=\frac{\partial^2 y}{\partial x_5^2}=\frac{\partial^2 y}{\partial y_B^2}=\frac{5}{27},\quad c_{66}=\frac{\partial^2 y}{\partial x_6^2}=\frac{\partial^2 y}{\partial z_B^2}=\frac{5}{27},$$

$$c_{12}=\frac{\partial^2 y}{\partial x_1\partial x_2}=\frac{\partial^2 y}{\partial x_A\partial x_A}=-\frac{2}{27},\quad c_{13}=\frac{\partial^2 y}{\partial x_1\partial x_3}=\frac{\partial^2 y}{\partial x_A\partial z_A}=-\frac{2}{27},$$

$$c_{14}=\frac{\partial^2 y}{\partial x_1\partial x_4}=\frac{\partial^2 y}{\partial x_A\partial x_B}=-\frac{8}{27},\quad c_{15}=\frac{\partial^2 y}{\partial x_1\partial x_5}=\frac{\partial^2 y}{\partial x_A\partial y_B}=\frac{2}{27};$$

$$c_{16}=\frac{\partial^2 y}{\partial x_1\partial x_6}=\frac{\partial^2 y}{\partial x_A\partial z_B}=\frac{2}{27},\quad c_{23}=\frac{\partial^2 y}{\partial x_2\partial x_3}=\frac{\partial^2 y}{\partial x_A\partial z_A}=-\frac{4}{27},$$

$$c_{24}=\frac{\partial^2 y}{\partial x_2\partial x_4}=\frac{\partial^2 y}{\partial y_A\partial x_B}=\frac{2}{27},\quad c_{25}=\frac{\partial^2 y}{\partial x_2\partial x_5}=\frac{\partial^2 y}{\partial x_A\partial z_B}=-\frac{5}{27},$$

$$c_{26}=\frac{\partial^2 y}{\partial x_2\partial x_6}=\frac{\partial^2 y}{\partial y_A\partial z_B}=\frac{4}{27},\quad c_{34}=\frac{\partial^2 y}{\partial x_3\partial x_4}=\frac{\partial^2 y}{\partial z_A\partial x_B}=\frac{2}{27},$$

$$c_{35}=\frac{\partial^2 y}{\partial x_3\partial x_5}=\frac{\partial^2 y}{\partial z_A\partial y_B}=\frac{4}{27},\quad c_{36}=\frac{\partial^2 y}{\partial x_3\partial x_6}=\frac{\partial^2 y}{\partial z_A\partial z_B}=-\frac{5}{27},$$

$$c_{45}=\frac{\partial^2 y}{\partial x_4\partial x_5}=\frac{\partial^2 y}{\partial z_B\partial y_B}=-\frac{2}{27},\quad c_{46}=\frac{\partial^2 y}{\partial x_4\partial x_6}=\frac{\partial^2 y}{\partial z_A\partial z_B}=-\frac{2}{27},$$

$$c_{56}=\frac{\partial^2 y}{\partial x_5\partial x_6}=\frac{\partial^2 y}{\partial y_B\partial z_B}=-\frac{4}{27}$$

由公式(4-3-5)得：

$$\begin{aligned}\sigma_{yn}^2=&\left(\frac{1}{3}\right)^2\frac{1}{2}+\left(\frac{2}{3}\right)^2\frac{1}{2}+\left(\frac{2}{3}\right)^2+\left(-\frac{1}{3}\right)^2\frac{1}{4}+\left(-\frac{2}{3}\right)^2\frac{1}{4}+\left(-\frac{2}{3}\right)^2\frac{1}{2}\\&+\frac{1}{2}\left[\left(\frac{8}{27}\right)^2\left(\frac{1}{2}\right)^2+\left(\frac{5}{27}\right)^2\left(\frac{1}{2}\right)^2+\left(\frac{5}{27}\right)^2+\left(\frac{8}{27}\right)^2\left(\frac{1}{4}\right)^2+\left(\frac{5}{27}\right)^2\left(\frac{1}{4}\right)^2+\left(\frac{5}{27}\right)^2\left(\frac{1}{2}\right)^2\right]\\&+\left(-\frac{2}{27}\right)^2\frac{1}{2}\,\frac{1}{2}+\left(-\frac{2}{27}\right)^2\frac{1}{2}+\left(-\frac{8}{27}\right)^2\frac{1}{2}\,\frac{1}{4}+\left(\frac{2}{27}\right)^2\frac{1}{2}\,\frac{1}{4}+\left(\frac{2}{27}\right)^2\frac{1}{2}\,\frac{1}{2}\\&+\left(-\frac{4}{27}\right)^2\frac{1}{2}+\left(\frac{2}{27}\right)^2\frac{1}{2}\,\frac{1}{4}+\left(-\frac{5}{27}\right)^2\frac{1}{2}\,\frac{1}{4}+\left(\frac{4}{27}\right)^2\frac{1}{2}\,\frac{1}{2}\\&+\left(\frac{2}{27}\right)^2\frac{1}{4}+\left(\frac{4}{27}\right)^2\frac{1}{4}+\left(-\frac{5}{27}\right)^2\frac{1}{2}\\&+\left(-\frac{2}{27}\right)^2\frac{1}{4}\,\frac{1}{4}+\left(-\frac{2}{27}\right)^2\frac{1}{4}\,\frac{1}{2}\\&+\left(-\frac{4}{27}\right)^2\frac{1}{4}\,\frac{1}{2}=1.0750\end{aligned}$$

若将 y 的展开式仅取至一次项，即按线性函数求 y 的方差，则有

$$\delta_{yl}^2=\left(\frac{1}{3}\right)^2\frac{1}{2}+\left(\frac{2}{3}\right)^2\frac{1}{2}+\left(\frac{2}{3}\right)^2+\left(-\frac{1}{3}\right)^2\frac{1}{4}+\left(-\frac{2}{3}\right)^2\frac{1}{4}+\left(-\frac{2}{3}\right)^2\frac{1}{2}=0.8056$$

可见 $\sigma_{yn}^2>\sigma_{yl}^2$. 这说明一个非线性函数若按线性函数计算其方差，则人为地将函数的精度拔高了.

二、未知数函数的方差

在测量数据处理中，用得最多的函数并非观测值的函数，而是未知数估值 $\hat{X}$ 的函数. 因为通过平差计算，求得了未知数的估值 $\hat{X}$ 后，便可根据它们来计算该平差问题中任一量的平差值，或者说该平差问题中的任一量都可以表达为未知数估值 $\hat{X}$ 的函数[19]（於宗俦、鲁林成，1983）. 那么，如何求得这些未知数估值 $\hat{X}$ 函数的方差呢？下面简要讨论这个问题.

1. 线性函数

设未知数估值 $\hat{X}$ 的线性函数为

$$Y=Y_0+K\hat{X} \tag{4-3-7}$$

式中：Y_0 为 $m\times 1$ 的常数向量；K 为 $m\times t$ 的常系数矩阵.

由(4-1-25)式知，未知数估值 $\hat{X}$ 的方差阵为

$$\mathrm{Var}(\hat{X})=\hat{\sigma}^2(B'B)^{-1}+\frac{1}{2}\hat{\sigma}^4M(2A_G+A_H)M'$$

由(4-3-7)式直接应用协方差传播律得：

$$Var(Y)=K\mathrm{Var}(\hat{X})K'=\hat{\sigma}^2K(B'B)^{-1}K'+\frac{1}{2}\hat{\sigma}^4KM(2A_G+A_H)M'K' \tag{4-3-8}$$

2.非线性函数

设未知数估值 $\hat{X}$ 的非线性函数为:

$$Y=\varphi_0+\varphi(\hat{X}) \tag{4-3-9}$$

式中:φ_0 为 $m\times 1$ 的常数向量.

将(4-3-9)式展为泰勒级数,取至一次项,并略去二次以上各项,得

$$Y=Y_0+K\Delta X$$

式中:

$$Y_0=\varphi_0+\varphi(X_0),K=\frac{\partial\varphi}{\partial\hat{X}}$$

令

$$\Delta Y=K\Delta X$$

则由(4-3-8)式可得未知数估值 $\hat{X}$ 的非线性函数 Y 的方差矩阵.

事实上推导未知数估值 $\hat{X}$ 的非线性函数的方差不应该线性近似.即将(4-3-9)式展为泰勒级数,至少应取至二次项,但这时公式推导十分复杂,还有待进一步研究.

第五章　非线性模型参数的其他估计

第一节　带约束的非线性模型参数估计

非线性模型(2-1-4)式隐含着 t 个未知数 $x_1, x_2, \cdots, x_t$ 相互独立这一假设.但在实际工作中,这一假设有时并不满足,即在实际工作中,往往所选取的参数个数 u 大于 t.于是就使得参数之间存在 $u-t=d$ 个约束条件.例如图5-1-1所示的测边网,A 点坐标 X_A, Y_A 已知;A 点至1点的方位角 α_{A1} 已知.现选取1、2、3点的坐标为参数,即:

$$X=(x_1 \quad y_1 \quad x_2 \quad y_2 \quad x_3 \quad y_3)'.$$

参数的个数 $u=6$,而本例的必要观测 $t=5$.于是 A 点坐标与1点坐标之间存在一个约束条件.于是有

$$\left.\begin{aligned}
L_1 &= \sqrt{(x_1-x_A)^2+(y_1-y_A)^2}+\Delta_1\\
L_2 &= \sqrt{(x_1-x_2)^2+(y_1-y_2)^2}+\Delta_2\\
L_3 &= \sqrt{(x_2-x_3)^2+(y_2-y_3)^2}+\Delta_3\\
L_4 &= \sqrt{(x_3-x_A)^2+(y_3-y_A)^2}+\Delta_4\\
L_5 &= \sqrt{(x_1-x_3)^2+(y_1-y_3)^2}+\Delta_5\\
L_6 &= \sqrt{(x_2-x_A)^2+(y_2-y_A)^2}+\Delta_6\\
&\arctan\frac{y_1-y_A}{x_1-x_A}=\alpha_{A_1}
\end{aligned}\right\} \qquad (5\text{-}1\text{-}1)$$

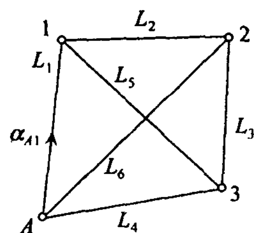


图5-1-1

式中前6个方程为观测方程,第7个方程是对参数 x_1, y_1 的约束,即 x_1, y_1 应满足的条件,称为约束方程或条件方程.(5-1-1)式就是带有约束的非线性模型.

一般地,带约束的非线性模型可写为

$$\left.\begin{aligned}
L &= f(X)+\Delta\\
\Phi(X) &= \alpha
\end{aligned}\right\} \qquad (5\text{-}1\text{-}2)$$

式中:$L=f(X)+\Delta$ 为观测方程(n 维向量);X 为 u 维未知参数;$\Phi(X)$ 为函数形式已知的 d 维向量,$d=u-t$;α 为已知的 d 维常数向量.

用估值代替真值,则(5-1-2)式可写为

$$\left.\begin{aligned}
V &= f(\hat{X})-L\\
\Phi(\hat{X})-\alpha &= 0
\end{aligned}\right\} \qquad (5\text{-}1\text{-}3)$$

为了求解未知参数 $\hat{X}$,组成新函数

$$\Psi = V'V+2K'(\Phi(\hat{X})-\alpha) \qquad (5\text{-}1\text{-}4)$$

式中:K 为d 维待定向量.

对(5-1-4)式求偏导数,并令其为0,再顾及(5-1-3)式,得

$$\left.\begin{aligned}&\left(\frac{\partial V}{\partial X}\right)'(f(\hat{X})-L)+\left(\frac{\partial \Phi(X)}{\partial X}\right)'K=0\\&\Phi(\hat{X})-\alpha=0\end{aligned}\right\}\tag{5-1-5}$$

求解非线性方程组(5-1-5)式,即可得到非线性模型(5-1-2)式的非线性最小二乘解.

由于函数 $f(X)$和 $\Phi(X)$的任意性都很大,一般求解(5-1-5)式并非易事.为此,我们在此只考虑一种最简单的、仅带有线性约束的情形.于是(5-1-3)式可写为

$$\left.\begin{aligned}&V=f(\hat{X})-L\\&C\hat{X}-\alpha=0\end{aligned}\right\}\tag{5-1-6}$$

d 个约束方程表明独立的未知参数只有 $u-d=t$ 个.故带约束的非线性模型(5-1-6)式等价于一个仅含 t 个独立未知参数的无约束非线性模型.只要找到这个等价的无约束非线性模型,带约束的非线性模型(5-1-6)式就可以变换为无约束非线性模型.下面就来寻找这个等价的无约束非线性模型.

从约束方程 $C\hat{X}-\alpha=0$ 解出未知参数 $\hat{X}$,得:

$$\hat{X}=C^{+}\alpha+F\hat{Y}=C'(CC')^{-1}\alpha+F\hat{Y}\tag{5-1-7}$$

式中:F 为$u\times t$ 的待定矩阵;$\hat{Y}$ 为$t\times 1$ 的新参数.F 由下式确定:

$$CF=0,\ \ F'F=I\tag{5-1-8}$$

将(5-1-7)式代入(5-1-6)式,得等价的无约束非线性模型为

$$V=g(\hat{Y})-L\tag{5-1-9}$$

其中

$$g(\hat{Y})=f(C'(CC')^{-1})\alpha+F\hat{Y}\tag{5-1-10}$$

由于新参数 $\hat{Y}$ 为t 维未知向量,故从无约束非线性模型(5-1-9)式,可用第三章介绍的任一种算法解出 $\hat{Y}$.将 $\hat{Y}$ 代入(5-1-7)式,就可得到带约束的非线性模型(5-1-6)式的全部参数 $\hat{X}$.

例 5-1-1 带约束的非线性为

$$\left.\begin{aligned}&L_i=ix_1^2+x_1x_2-e^{ix_3}+\Delta_i\quad(i=1,2,\cdots,n)\\&-x_1+x_2-x_3-3.24=0\end{aligned}\right\}\tag{5-1-11}$$

设 x_1,x_2 的真值为 $x_1=1.85$, $x_2=2.34$.用此真值算出的观测值真值 $\tilde{L}_i$ 及随机观测值 L_i 列于表 5-1-1.试用表 5-1-1 中的观测值估计参数 $\hat{X}=(\hat{x}_1,\hat{x}_2,\hat{x}_3)'$.

表 5-1-1

i	1	2	3	4	5	6
真值	7.68757214	11.16991323	14.59623874	18.0189833	21.44149893	24.86399993
观测值	7.649	11.261	14.566	17.969	21.520	24.945

解:由(5-1-11)式知,$C=(-1\quad 1\quad -1)$,$\alpha=3.24$,设 $F'=\begin{pmatrix}f_{11}&f_{21}&f_{31}\\f_{12}&f_{22}&f_{32}\end{pmatrix}$.由(5-1-8)

式得

$$C'(CC')^{-1}\alpha = (-1.08 \quad 1.08 \quad -1.08)$$

$$-f_{11}+f_{21}-f_{31}=0$$

$$-f_{12}+f_{22}-f_{32}=0$$

$$f_{11}^2+f_{21}^2+f_{31}^2=1$$

$$f_{12}^2+f_{22}^2+f_{32}^2=1$$

$$f_{11}f_{12}+f_{21}f_{22}+f_{31}f_{32}=0$$

令 $f_{11}=0$，可解得 $F'=\begin{pmatrix} 0 & 0.707106 & 0.707106 \\ -0.816496 & -0.408248 & 0.408248 \end{pmatrix}$，于是

$$\hat{X}=\begin{pmatrix}\hat{x}_1\\ \hat{x}_2\\ \hat{x}_3\end{pmatrix}=\begin{pmatrix}-1.08\\ 1.080\\ -1.08\end{pmatrix}+\begin{pmatrix}-0.816496\hat{y}_2\\ 0.707106\hat{y}_1-0.408248\hat{y}_2\\ 0.707106\hat{y}_1+0.408248\hat{y}_2\end{pmatrix}=\begin{pmatrix}-0.816496\hat{y}_2-1.08\\ 0.707106\hat{y}_1-0.408248\hat{y}_2+1.08\\ 0.707106\hat{y}_1+0.408248\hat{y}_2-1.08\end{pmatrix}$$

从而可得与(5-1-11)式等价的无约束非线性模型为

$$V_i = i(-0.816496\hat{y}_2-1.08)^2+(-0.816496\hat{y}_2-1.08)(0.707106\hat{y}_1-0.408248\hat{y}_2 +1.08)-e^{i(0.707106\hat{y}_1+0.408248\hat{y}_2-1.08)}-L_i,\ i=1,2,\cdots,6$$

根据上式，应用单纯形法迭代 68 次，得

$$\hat{y}_1=-0.341918,\quad \hat{y}_2=-3.592283903$$

从而解得

$$\hat{X}=(1.852962 \quad 2.306129 \quad -2.786833)',\qquad \| X-\hat{X}\|_2=0.050$$

第二节　非线性模型参数的稳健估计

在第一章第一节中，我们简要回顾了线性模型的抗差最小二乘估计. 在线性模型的抗差最小二乘估计中，对非线性模型总是进行线性近似. 因为线性近似不可避免地会引起模型误差，用具有模型误差的近似模型去处理粗差，可能会造成错误的粗差鉴别和定位[7]（周江文等，1997）. 因此，要解决这个问题，最好是寻求一种适合于非线性模型的稳健估计方法. 由于这方面的研究成果目前还不多见，故本节仅给出一种简单实用的非线性抗差最小二乘解.

设不同精度独立观测值为 $L=(L_1\ L_2\cdots\ L_n)'$，相应的权矩阵为 $P=\mathrm{diag}(p_1\ p_2\cdots\ p_n)$，其非线性误差方程为：

$$V=f(\hat{X})-L \tag{5-2-1}$$

(5-2-1)式的纯量形式为：

$$v_i=f_i(\hat{X})-L_i \tag{5-2-2}$$

相应的权为 p_i.

求解误差方程(5-2-1)式的准则函数可表达为下列最优化问题：

$$\sum p_i\rho(v_i)=\min \tag{5-2-3}$$

式中最优化准则函数 $\rho(v_i)$ 取不同的形式，可得到不同的最优化准则.

将(5-2-3)式对 x_i 求偏导数，并令其为零，得

$$\sum p_i \frac{\partial \rho(v_i)}{\partial v_i} \frac{\partial v_i}{\partial x_i} = 0 \tag{5-2-4}$$

令

$$\left.\begin{aligned} \frac{\partial \rho(v_i)}{\partial v_i} &= \Psi(v_i) \\ w_i &= \frac{\Psi(v_i)}{v_i} \end{aligned}\right\} \tag{5-2-5}$$

将(5-2-5)式代入(5-2-4)式得

$$\sum p_i w_i v_i \frac{\partial v_i}{\partial x_i} = 0 \tag{5-2-6}$$

令

$$\overline{P}_i = p_i w_i \tag{5-2-7}$$

式中：$\overline{P}_i$ 称为等价权；w_i 称为权因子.

将(5-2-7)式代入(5-2-6)式，并顾及(5-2-2)式，得

$$\sum \overline{P}_i \frac{\partial f_i(X)}{\partial x_i}(f_i(X) - L_i) = 0 \tag{5-2-8}$$

其矩阵形式为

$$\left(\frac{\partial f_i(X)}{\partial x_i}\right)' \overline{P}(f(\hat{X}) - L_i) = 0 \tag{5-2-9}$$

当 $f(\hat{X})$ 为线性函数时，(5-2-9)式就是线性模型抗差最小二乘估计的方法. 现在 $f(\hat{X})$ 为非线性函数，故称满足(5-2-9)式的解为非线性抗差最小二乘解.

由于 $f(\hat{X})$ 为非线性函数，故(5-2-9)式没有显表达式. 求解时可应用第三章介绍的任一种方法.

在求解(5-2-9)式之前，还必须选择最优化准则函数 $\rho(V)$ 或 $\Psi(V)$. 关于 $\Psi(V)$ 的选择，李德仁院士给出了五条选择原则[59]（李德仁，1988）.

根据这五条原则，现已发展了很多种 $\Psi(V)$ 函数. 最近，J. Saleh 又从能量最小原理出发，导出了各种情况下基于能量最小的权因子表达式[60]（J. Saleh，2000）. 在这诸多权因子的选择中，周江文提出的 IGG 法权因子是值得推荐的一种[61]（周江文，1989）. 周江文建议的权因子可写为

$$w(v_i) = \begin{cases} 1, & |v_i| \leqslant 1.5\sigma \\ \dfrac{k\sigma}{|v_i|}, & 1.5\sigma < |v_i| \leqslant 2.5\sigma \\ 0, & |v_i| > 2.5\sigma \end{cases} \tag{5-2-10}$$

例 5-2-1 设例2-1-1中含有粗差，其观测值为 $L_3 = 8.52$，其余观测值见表 2-1-1. 试求参数的估值 $\hat{X}$.

直接按例 3-3-5 中的高斯-牛顿法程序求解，迭代 16 次，得

$$\hat{X} = (5.727745804 \quad -0.133524000)', \ \| X - \hat{X} \|_2 = 0.3305$$

而由例 3-3-5 知，当无粗差存在时，其解为

$$\hat{X} = (5.422744573 \quad -0.255672086)', \ \| X - \hat{X} \|_2 = 0.0029$$

可见无粗差影响时解的精度比有粗差影响时要高两个数量级. 为了抵抗粗差对参数估计的影响，下面采用非线性抗差最小二乘估计求解. 先按(5-2-10)式确定权因子，然后仍然采用高斯-牛顿法求解，迭代结果列于表 5-2-1. 经过 10 次迭代得

$$\hat{X} = (5.422978965 \quad -0.255760000)', \ \| X - \hat{X} \|_2 = 0.0032$$

与无粗差影响时的结果几乎相等.

表 5-2-1

k	1	2	3	4	5	6	7	8	9	10
$\hat{X}_1^{(k)}$	4.2308	4.9567	5.2518	5.39189	5.42223	5.42296	5.42297	5.42297	5.42297	5.422978965
$\hat{X}_2^{(k)}$	0.0601	−0.1137	−0.2088	−0.25001	−0.25559	−0.25575	−0.25575	−0.25576	−0.25576	−0.25576000
v_1	−0.8250	0.5533	−0.4363	−0.88987	−1.05051	−1.07414	−1.07479	−1.07481	−1.07481	−1.07481500
v_2	−0.5715	1.2556	−0.7153	−1.42847	−1.63308	−1.66091	−1.66166	−1.66169	−1.66169	−1.66169400
v_3	−6.4036	−4.3110	−6.9937	7.74355	7.91172	−7.93347	−7.93404	−7.93406	−7.93407	−7.93407000
v_4	−0.2843	1.9081	−1.2446	−1.88941	−1.98324	−1.99462	−1.99491	−1.99491	−1.99491	−1.99491800
v_5	−0.2033	1.9376	−1.4667	−1.92302	−1.93070	−1.93091	−1.93088	−1.93087	−1.93087	−1.93087700
p_1	1	1	1	1	1	1	1	1	1	1
p_2	1	1	1	1	1	1	1	1	1	1
p_3	0.3425	0.0961	0.0059	3×10^{-4}	2×10^{-5}	1×10^{-6}	8×10^{-8}	4×10^{-9}	3×10^{-10}	3×10^{-10}
p_4	1	1	1	1	1	1	1	1	1	1
p_5	1	1	1	1	1	1	1	1	1	1
$\hat{\sigma}$	3.74770	2.28350	1.74190	1.86706	1.95389	1.96813	1.96844	1.96845	1.96845	1.96845

第三节　非线性模型参数的拟似然估计

将(2-1-4)式的随机模型一起写出，有

$$\begin{aligned} L &= f(X) + \Delta \\ E(\Delta) &= 0 \\ \mathrm{Var}(\Delta) &= \mathrm{Var}(L) = \sigma^2 Q_{LL} = \sigma^2 P^{-1} \end{aligned} \tag{5-3-1}$$

一、拟似然估计

设观测值 L 服从一类概率分布，其概率密度为 $p(X)$，其对数似然函数为 $l(X)$. 概率密度 $p(X)$关于 X 存在一阶偏导数. 我们定义得分函数为

$$S(X) = \frac{\partial l(X)}{\partial X} \tag{5-3-2}$$

如果 L 的分布为自然指数族分布，则

$$S(X) = B'(X)PV(X) \tag{5-3-3}$$

式中：

$$B(X) = \frac{\partial f(X)}{\partial X}, \ V(X) = f(X) - L \tag{5-3-4}$$

如果 L 的分布类型未知,则假定得分函数仍与(5-3-3)式同形,即此时的得分函数为

$$q(X)=B'(X)PV(X) \tag{5-3-5}$$

(5-3-5)式称为模型(5-3-1)式的拟得分函数.

当拟得分函数(5-3-5)式等于0时,称为拟似然方程,即拟似然方程为

$$q(X)=B'(X)PV(X)=0 \tag{5-3-6}$$

使拟似然方程(5-3-6)式成立得解 $\hat{X}^*$ 称为未知参数 X 的拟似然估计[62](林路,1999).

将(5-3-4)式代入(5-3-6)式,则拟似然方程可写为

$$\left(\frac{\partial f(X)}{\partial X}\right)'Pf(X)-\left(\frac{\partial f(X)}{\partial X}\right)'PL=0 \tag{5-3-7}$$

所以非线性模型(5-3-1)式的拟似然估计就是非线性方程组(5-3-7)式的解 $\hat{X}^*$. 关于解算非线性方程组,已有很多好的算法. 附录C所摘录的文章[63](陈志、高旅端、邓乃扬,1998)就是一种解算非线性方程组的很好的算法.

二、泛拟似然估计

与拟似然估计类似,泛拟似然估计是使下列方程

$$G(X)=B'(X)C'(X)V(X)=0 \tag{5-3-8}$$

成立的解 $\hat{X}$.

式中:

$$C=A'(AQ_{LL}A')^{-}A \tag{5-3-9}$$

或

$$C=A'(AQ_{LL}A')^{-}AB(X)B^{-}(X) \tag{5-3-10}$$

A 为任意 $k\times n$ 阶矩阵.

三、拟似然估计的性质

一切形如 $G(X)$ 的函数所构成的函数类 F,称为广义正则线性无偏函数类. F 中的函数有无穷多个,且 $q(X)\in F$. 在函数类 F 的无穷多个函数中任取一函数 $G(X)\in F$,我们定义 $G(X)$ 与 $S(X)$ 的距离为[62](林路,1999)

$$d(G,S)=\mathrm{E}((S(X)-G(X))(S(X)-G(X))') \tag{5-3-11}$$

性质1. 在此距离下,$q(X)$ 是函数类 F 的无穷多个函数中与 $S(X)$ 的距离最近的一个.

性质2. 拟似然估计是泛拟似然估计类中的渐近最优估计.

以上性质的详细证明见文献[62](林路,1999).

第四节　非线性模型参数的贝叶斯估计

在数理统计领域,存在两大学派——经典学派(或频率学派)和贝叶斯(Bayes, T. R.)学派. "经典学派"也称为抽样学派,是指20世纪初由英国的卡尔·皮尔逊(Pearson, K.)等人开始,经费歇(Fisher, R. A.)予以发展,到奈曼(Neyman, J.)完成理论的这一系统的成果. 在目前国内已出版的教材中,这一部分——指经典学派的方法和理论,往往占有全部或绝大部分的篇幅,如点估计、假设检验、矩估计法、最大似然估计法、最小二乘法、估计的优良性质、区间估计、假设检验中犯两类错误的概率、优良的假设检验方法等等[64](张尧庭,陈汉峰,

1991).“贝叶斯学派奠基性的工作是贝叶斯的论文(Bayes, T. R., 1763). 可能是他自己感到他的学说还有不完善的地方,这一论文在他生前没有发表,而是在他死后由贝叶斯的朋友发表的. 著名的数学家拉普拉斯(Laplace, P. S.)用贝叶斯提出的方法,导出了重要的‘相继律’后,贝叶斯的方法和理论才逐渐被人理解和重视起来”[64](张尧庭,陈汉峰,1991). 现在,贝叶斯学派已成长为统计学中一个很有影响和很有力量的学派,其势头还在不断增长.“已经可以肯定地说,贝叶斯学派已经形成并发展为一个在统计学中有影响的、堪与频率学派并列的学派”[65](陈希孺,1990). 而“Lindley 认为,21 世纪将是贝叶斯统计的世界”[66](成平,1990).

“贝叶斯学派的起点是贝叶斯的两项工作:贝叶斯定理和贝叶斯假设”[64](张尧庭,陈汉峰,1991).

1. 贝叶斯定理

定理 5-4-1 设 $L^T=(l_1 \quad l_2 \quad \cdots \quad l_n)$为观测向量,其分布密度为 $f(L/X)$. 这里 $X^T=(x_1 \quad x_2 \quad \cdots \quad x_t)$, $X\in\Theta$, 且 X 具有先验分布密度 $\pi(X)$. 于是, X 的后验分布密度为

$$f(X/L)\propto f(L/X)\pi(X) \tag{5-4-1}$$

式中: $\propto$表示成比例.

证明: 由条件密度函数 $f(L/X)=\dfrac{f(L,X)}{\pi(X)}$可得

$$f(L,X)=f(L/X)\pi(X)=f(L)f(X/L) \tag{5-4-2}$$

式中: $f(L,X)$称为 L 和 X 的联合分布密度函数.

由(5-4-2)式可得

$$f(X/L)=\frac{f(L/X)\pi(X)}{f(L)} \tag{5-4-3}$$

在观测值 L 给定的条件下, $f(L)$是与 X 无关的函数. 所以 $f(X/L)$与 $\pi(X)f(L/X)$成比例. 其比例常数为$\dfrac{1}{f(L)}$. 定理得证[67](王新洲,1994).

因为 $$\int_\Theta f(X/L)\mathrm{d}x = \int_\Theta f(L/X)\pi(X)\mathrm{d}x/f(L) = 1$$

所以 $$f(L)=\int_\Theta f(L/X)\pi(X)\mathrm{d}x \triangleq 1/C \tag{5-4-4}$$

于是,贝叶斯定理还可以写为

$$f(X/L)=Cf(L/X)\pi(X) \tag{5-4-5}$$

式中: $$C=\frac{1}{\int_\Theta f(L/X)\pi(X)\mathrm{d}x} \tag{5-4-6}$$

称为非标准常数.

先验分布密度 $\pi(X)$概括了试验者在试验之前对未知参数 X 的认识,而 $f(X/L)$则是在得到样本 L 的条件下,试验者对 X 的重新认识,称为 X 的后验分布密度. 后验分布密度综合了 X 的先验信息和样本带来的关于 X 的新信息. 先验信息蕴含在先验分布密度 $\pi(X)$之中,样本带来的信息蕴含在条件密度 $f(L/X)$之中. 如果把 $f(L/X)$视为 X 的函数,则称为似然函数,记为 $L(X/L)$. 于是后验分布密度可写为:

$$f(X/L)\propto L(X/L)\pi(X) \tag{5-4-7}$$

或
$$f(X/L)=CL(X/L)\pi(X) \tag{5-4-8}$$

(5-4-7)式表明:在得到样本之后,X 的后验分布密度与似然函数和先验分布密度的乘积成正比.

2. 贝叶斯假设

由贝叶斯定理知,要得到后验分布密度 $f(X/L)$,必须已知 X 的先验分布密度 $\pi(X)$;然而,在实际工作中,有时对参数 X 是没有任何过去的知识可以借鉴的.这种情况下的先验分布就称为无信息先验分布.由于对 X 一无所知,所以贝叶斯假设参数 X 的无信息先验分布密度 $\pi(X)$ 应在其取值范围内是"均匀"分布的.又因为对 X 无任何信息可以借鉴,故可设 X 的取值范围为 $(-\infty,\infty)$.于是,根据广义分布密度[64](张尧庭,陈汉峰,1991)知:

$$\pi(X)=C,\quad -\infty<X<\infty$$

或
$$\pi(X)=1,\quad -\infty<X<\infty \tag{5-4-9}$$

将(5-4-9)式代入(5-4-7)式,得

$$f(X/L)\propto L(X/L) \tag{5-4-10}$$

(5-4-10)式表明:当采用无信息先验分布时,似然函数是后验分布密度的核[64](张尧庭,陈汉峰,1991).

3. 贝叶斯估计

(1)损失函数与风险函数

评定估计量好坏的标准,通常用损失函数和风险函数来描述.

在参数空间 Θ 上定义的一个二元非负实函数 $S(X,\hat{X})$ 称为损失函数.$S(X,\hat{X})$ 表示用 $\hat{X}$ 去估计 X 时,由于 $\hat{X}$ 与 X 不同而引起的损失.一般地,损失总是非负的.因此,限定 $S(X,\hat{X})\geqslant 0$.据此定义,可得到很多种形式的损失函数.在参数估计中,最常用的损失函数为平方损失(或二次损失)函数:

$$S(X,\hat{X})=(X-\hat{X})'P(X-\hat{X}) \tag{5-4-11}$$

对于损失函数 $S(X,\hat{X})$,当用 $\hat{X}$ 去估计 X 时,其平均损失

$$R(X,\hat{X})=\mathrm{E}(S(X,\hat{X})) \tag{5-4-12}$$

称为风险函数.

(2)贝叶斯估计

当我们取二次损失函数(5-4-11)式作为损失函数时,其风险函数为

$$\begin{aligned}S(X,\hat{X})&=\mathrm{E}(S(X,\hat{X}))=\mathrm{E}((X-\hat{X})'P(X-\hat{X}))\\&=\mathrm{E}((X-\mathrm{E}(X)+\mathrm{E}(X)-\hat{X})'P(X-\mathrm{E}(X)+\mathrm{E}(X)-\hat{X}))\end{aligned} \tag{5-4-13}$$

由于 $\mathrm{E}(X-\mathrm{E}(X))=0$,所以(5-4-13)式变为

$$S(X,\hat{X})=\mathrm{E}((\mathrm{E}(X)-\hat{X})'P(\mathrm{E}(X)-\hat{X})) \tag{5-4-14}$$

损失函数选定后,一个好的估计应使风险函数尽可能地小.在(5-4-14)式中,因为 P 为正定矩阵,故当 $\hat{X}=\mathrm{E}(X)$ 时,(5-4-14)式达到最小.于是在先验分布密度 $\pi(X)$ 下,以二次损失为损失函数的贝叶斯估计为

$$\hat{X}_B=\mathrm{E}(X) \tag{5-4-15}$$

顾及数学期望 $\mathrm{E}(X)$ 的定义,(5-4-15)式可写为

$$\hat{X}_B = \int_{A} Xf(X/L)\mathrm{d}X \qquad (5\text{-}4\text{-}16)$$

式中：$f(X/L)$是由(5-4-3)式或(5-4-8)式或(5-4-10)式所确定的 X 的后验密度函数. 显然，(5-4-16)式是非线性方程，且一般情况下(5-4-16)式不能用解析法计算积分，只能采用数值积分. 附录B所介绍的蒙特卡罗积分是较合适的数值积分法[67]（王新洲，1994）.

第五节 非线性模型非参数小波估计

估计理论中存在参数估计和非参数估计两类估计问题. 前面已详细讨论了非线性模型下的参数估计问题. 这节将扼要介绍非线性模型下的非参数估计问题.

设非线性非参数估计模型为

$$L = g(t) + \Delta \qquad (5\text{-}5\text{-}1)$$

式中：$L=(l_1 \quad l_2 \quad \cdots \quad l_n)'$为 $n\times1$ 阶的独立观测向量；$t=(t_1 \quad t_2 \quad \cdots \quad t_n)'$服从$[0,1]$上的均匀分布，$t_i=\frac{i}{n}\in[0,1]$；$\Delta$ 服从$N(0,\mathrm{Var}(L))$；$g(t)$是定义在$[0,1]$上的未知光滑函数.

Donoho 等人对模型(5-5-1)式提出了非线性小波估计[68]（张双林，沙秋英，程美玉，1999）. 现扼要介绍如下.

设 $\varphi(t)$和 $\psi(t)$是尺度函数和母小波，要求 $\varphi(t)$和 $\psi(t)$紧支撑，有连续导数，且$\psi(t)$有 $N-1$ 阶消失矩，即 $\psi(t)$和任意 $N-1$ 阶多项式正交，则对任何正整数 j_0 满足 $2^{j_0}\geqslant N$，可找到函数集合

$$\{\varphi_{j_0k}(t), k=0,1,\cdots,2^{j_0}-1; \psi_{jk}(t), j\geqslant j_0, k=0,1,\cdots,2^j-1\}$$

构成 $L^2[0,1]$标准正交基. 于是对任意 $f\in L^2[0,1]$有下列在 $L^2[0,1]$意义下收敛的展开式：

$$f(t) = \sum_{k=0}^{2^{j_0}-1} a_{j_0k}\varphi_{j_0k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} b_{jk}\psi_{jk}(t) \qquad (5\text{-}5\text{-}2)$$

式中：

$$a_{j_0k} = \int_0^1 f(t)\varphi_{j_0k}(t)\mathrm{d}t, \quad b_{jk} = \int_0^1 f(t)\psi_{jk}(t)\mathrm{d}t \qquad (5\text{-}5\text{-}3)$$

这样得到的 $L^2[0,1]$上的一组标准正交基也继承了尺度函数 $\varphi(t)$和母小波 $\psi(t)$的性质，即紧支撑，有一阶连续导数，并且对固定的 j_0，当 $k=N,N+1,\cdots,2^{j_0}-N-1$ 时有

$$\varphi_{j_0k}(t) = 2^{j_0/2}\varphi(2^{j_0}t-k) \qquad (5\text{-}5\text{-}4)$$

对固定的 j，当 $k=N,N+1,\cdots,2^j-N-1$ 时有

$$\psi_{jk}(t) = 2^{j/2}\varphi(2^jt-k) \qquad (5\text{-}5\text{-}5)$$

对与边界0和1有关的 $\varphi(t)$和 $\psi(t)$作适当的修正.

于是，对于非线性非参数估计模型(5-5-1)式，Hall 和 Patil[69]（1993）给出的非线性小波估计为

$$\hat{g}(t) = \sum_{k=0}^{2^{j_0}-1} \hat{a}_{j_0k}\varphi_{j_0k}(t) + \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} \hat{b}_{jk}I_{\{|\hat{b}_{jk}|>\delta\}}\psi_{jk}(t) \qquad (5\text{-}5\text{-}6)$$

式中：
$$\hat{a}_{j_0k}=\frac{1}{n}\sum_{i=1}^{n}l_i\varphi_{j_0k}(t_i),\hat{b}_{jk}=\frac{1}{n}\sum_{i=1}^{n}l_i\psi_{jk}(t_i)\tag{5-5-7}$$

张双林，沙秋英，程美玉(1999)[68]给出的非线性小波估计为

$$\hat{g}(t)=\sum_{k=0}^{2^{j_0}-1}\hat{a}_{j_0k}\varphi_{j_0k}(t)+\sum_{j=j_0}^{J_1}\sum_{k=0}^{2^j-1}\hat{b}_{jk}I_{\{|\hat{b}_{jk}|>\delta\}}\psi_{jk}(t)\tag{5-5-8}$$

式中：
$$\begin{aligned}\hat{a}_{j_0k}&=\frac{1}{n}\sum_{i=1}^{n}l_i\varphi_{j_0k}(t_i)I_{\{|l_i\varphi_{j_0k}(t_i)|\leqslant M\}}\\ \hat{b}_{jk}&=\frac{1}{n}\sum_{i=1}^{n}l_i\psi_{jk}(t_i)I_{\{|l_i\psi_{jk}(t_i)|\leqslant M\}}\\ M&=\sqrt{n}\end{aligned}\tag{5-5-9}$$

可以证明，由(5-5-8)式和(5-5-9)式得到的$\hat{g}(t)$是$g(t)$的一致强相合估计. 这一结论的证明见张双林，沙秋英，程美玉所著《回归函数非线性小波估计的一致强相合性》[68](1999).

第六节　非线性模型半参数估计

将非线性参数估计模型(2-1-4)式与非线性非参数估计模型(5-5-1)式综合在一起，就可以得到如下的新模型：

$$\begin{aligned}&L=f(X)+g(t)+\Delta\\&\mathrm{E}(\Delta)=0\\&\mathrm{Var}(\Delta)=\mathrm{Var}(L)=\sigma^2P^{-1}\end{aligned}\tag{5-6-1}$$

式中：L 为独立观测向量；X 为未知参数向量；$f(X)$为函数形式已知的非线性函数；$t=(t_1\quad t_2\quad\cdots\quad t_n)'$服从$[0,1]$上的均匀分布，$t_i=\frac{i}{n}\in[0,1]$；$\Delta$ 服从$N(0,\mathrm{Var}(L))$；$g(t)$是定义在$[0,1]$上的未知光滑函数.

模型(5-6-1)式称为非线性半参数估计模型. 非线性半参数估计模型包含了非线性参数估计模型和非线性非参数估计模型，即当$g(t)=0$时，(5-6-1)式变为(2-1-4)式，当$f(X)=0$时，(5-6-1)式变为(5-5-1)式.

模型(5-6-1)式的对数似然函数可写为

$$d_i(X,g(t_i))=-\frac{1}{2}\lg(2\pi\sigma^2)-\frac{1}{2\sigma^2}(l_i-f(X)-g(t_i))^2\quad(1\leqslant i\leqslant n)\tag{5-6-2}$$

令

$$D_n(X,g)=\sum_{i=1}^{n}d_i(X,g(t_i))\tag{5-6-3}$$

下面根据(5-6-2)式与(5-6-3)式来估计未知参数 X 和未知光滑函数$g(t)$.

为了估计未知参数 X，可将非参数成分视为无穷维多余参数. 为了叙述方便，我们以$\tilde{X}$和$\tilde{g}(t)$表示 X 和$g(t)$的真值. 假定存在一条曲线 $X\mapsto g(X,t)$满足 $g(\tilde{X},t)=\tilde{g}(t)$，且$\left.\frac{\partial g(X,t)}{\partial X}\right|_{X=\tilde{X}}=g^*(t)$，当 n 充分大时，下式成立：

$$\mathrm{E}\left\{\sum_{i=1}^{n}\left(\frac{\partial d_i(X,g(X,t_i))}{\partial X}\right)\left(\frac{\partial d_i(X,g(X,t_i))}{\partial g}\right)g(t_i)\right\}\bigg|_{X=\tilde{X}} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\frac{\partial f(\tilde{X})}{\partial X}+g^*(t_i)\right)g(t_i) = O(n^{\frac{1}{2}}) \tag{5-6-4}$$

式中：$g(t_i)$为任一定义在[0,1]上的光滑曲线. 曲线 $g(X,t)$称为渐近有效曲线，$g^*(t)$称为渐近有效方向[70]（朱仲义、韦博成，2001）.

如果非参数 g 是一个有限维多余参数，那么渐近有效曲线就是参数正交化变换. 由于渐近有效曲线 $g(X,t)$未知，假如能求得 $g(X,t)$的相合估计 $\hat{g}(X,t)$，则把 $D_n(X,\hat{g}(X))$称为广义子集参数似然函数. 下面根据广义子集参数似然函数 $D_n(X,\hat{g}(X))$来估计未知参数 X 和光滑曲线$g(t)$.

令 $W_{ni}(t)=W_{ni}(t_1 \quad t_2 \quad \cdots \quad t_n)$是一列仅依赖 t 的概率权函数，且 $\hat{g}(X,t)$是下列方程的解：

$$\sum_{i=1}^{n} W_{ni}(t)\frac{\partial d_i(X,g)}{\partial g} = \frac{1}{\sigma^2}\sum_{i=1}^{n} W_{ni}(t)(l_i - f_i(X)) - g = 0 \tag{5-6-5}$$

由(5-6-5)式可以求得 $\hat{g}(X,t)$的一个显式解：

$$\hat{g}(X,t) = \sum_{i=1}^{n} W_{ni}(t)(l_i - f_i(X)) \tag{5-6-6}$$

可以证明 $\hat{g}(X,t)$是 $g(X,t)$的相合估计[70]（朱仲义、韦博成，2001）.

再令

$$h(X) = (h_1(X) \quad h_2(X) \quad \cdots \quad h_n(X))'$$

$$h_i(X) = f_i(X) - \sum_{i=1}^{n} W_{ni}(t)(l_i - f_i(X))$$

$$B'(X) = \frac{\partial h(X)}{\partial X} \tag{5-6-7}$$

$$V(X) = h(X) - L \tag{5-6-8}$$

于是基于广义子集参数似然函数 $D_n(X,\hat{g}(X))$的极大似然估计 $\hat{X}$ 应满足下列方程：

$$B'(\hat{X})PV(\hat{X}) = \left(\frac{\partial h(\hat{X})}{\partial X}\right)' Ph(\hat{X}) - \left(\frac{\partial h(\hat{X})}{\partial X}\right)' PL = 0 \tag{5-6-9}$$

解非线性方程组(5-6-9)式，即可得到未知参数的极大似然估计 $\hat{X}$. 将 $\hat{X}$ 代入(5-6-6)式即可得到光滑曲线 $g(t)$的估值 $\hat{g}(t)$.

其实，(5-6-8)式就是与(3-1-1)式同型的非线性误差方程。故采用第三章介绍的任何一种方法都能得到(5-6-8)式中参数 X 的非线性最小二乘估计。

第六章　非线性模型参数估计在测量上的应用

第一节　在导线网平差中的应用

传统的平面控制网，如测角网、测边网、边角网已被 GPS 所取代．只有导线网在城市测量中还能发挥一些作用．故在此只简要叙述非线性模型参数估计在导线网平差中的应用．

我们知道，在导线测量中，其基本观测值为归算到高斯投影平面上的边长和方向，且边长观测值大约为 1/3，方向观测值大约为 2/3[71]（吴俊昶，刘大杰，于正林，1998）．由《控制网测量平差》（吴俊昶，刘大杰，于正林，1998）知，对于如图 6-1-1 所示的边长观测值 S_{ij}，其误差方程为

$$v_{S_{ij}} = \sqrt{(\hat{x}_j - \hat{x}_i)^2 + (\hat{y}_j - \hat{y}_i)^2} - S_{ij} \tag{6-1-1}$$

而方向观测值，如图 6-1-2 中的 β_{ij} 和 β_{ik}，其误差方程为

$$\begin{aligned} v_{\beta_{ij}} &= -\hat{Z}_i + \arctan\frac{\hat{y}_j - \hat{y}_i}{\hat{x}_j - \hat{x}_i} - \beta_{ij} \\ v_{\beta_{ik}} &= -\hat{Z}_i + \arctan\frac{\hat{y}_k - \hat{y}_i}{\hat{x}_k - \hat{x}_i} - \beta_{ik} \end{aligned} \tag{6-1-2}$$

式中：$\hat{Z}_i$ 为测站 i 点定向角未知数，即从纵坐标方向顺时针旋转到度盘零位置的角度[71]（吴俊昶，刘大杰，于正林，1998）．

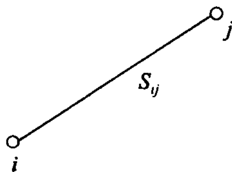


图 6-1-1

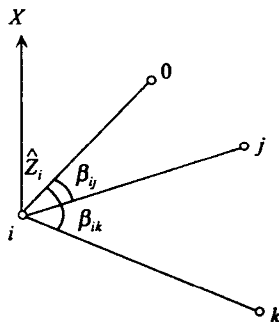


图 6-1-2

0 方向本身的误差方程为

$$v_{\beta_{i0}}=-\hat{Z}_i+\arctan\frac{\hat{y}_0-\hat{y}_i}{\hat{x}_0-\hat{x}_i}-\beta_{i0} \tag{6-1-3}$$

因为定向角未知数 $\hat{Z}_i$ 为测站点 i 至零方向 0 点的方位角的估值，而测站点 i 至零方向 0 点的方位角的估值可由这两点的坐标估值表示为

$$\hat{Z}_i=\arctan\frac{\hat{y}_0-\hat{y}_i}{\hat{x}_0-\hat{x}_i} \tag{6-1-4}$$

将(6-1-4)式代入(6-1-2)式和(6-1-3)式，得如图 6-1-2 中 β_{ij}、β_{ik}和 0 方向 β_{i0}的误差方程为

$$\begin{aligned}
v_{\beta_{i0}}&=0\\
v_{\beta_{ij}}&=\arctan\frac{\hat{y}_j-\hat{y}_i}{\hat{x}_j-\hat{x}_i}-\arctan\frac{\hat{y}_0-\hat{y}_i}{\hat{x}_0-\hat{x}_i}-\beta_{ij}\\
v_{\beta_{ik}}&=\arctan\frac{\hat{y}_k-\hat{y}_i}{\hat{x}_k-\hat{x}_i}-\arctan\frac{\hat{y}_0-\hat{y}_i}{\hat{x}_0-\hat{x}_i}-\beta_{ik}
\end{aligned} \tag{6-1-5}$$

(6-1-5)式为消去定向角未知数后的方向观测值的误差方程. 由于 0 方向的误差方程恒等于 0, 故任一测站上方向误差方程的个数为 n_i-1(n_i 为测站 i 上的方向数).

对于导线网, 分别按(6-1-1)式和(6-1-4)式列出误差方程后, 在 $V'PV=\min$ 下, 用第三章的任一种解算方法, 都能解算出未知参数的非线性最小二乘估计 $\hat{X}$.

例 6-1-1 导线网如图6-1-3所示, A、B 两点的坐标已知, 且无误差. 方向观测值的中误差为 σ_β, 各边长观测值的中误差分别为 σ_{S_i}. 试用非线性最小二乘估计平差该导线网.

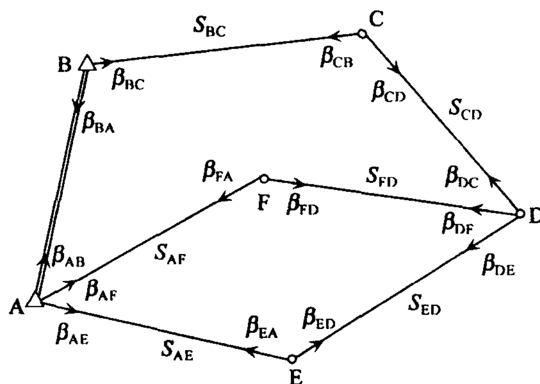


图 6-1-3

根据(6-1-1)式和(6-1-2)式列出边长观测值的误差方程, 即

$$\begin{aligned}
v_{S_{bc}}&=\sqrt{(\hat{x}_c-x_b)^2+(\hat{y}_c-y_b)^2}-S_{bc}\\
v_{S_{cd}}&=\sqrt{(\hat{x}_d-\hat{x}_c)^2+(\hat{y}_d-\hat{y}_c)^2}-S_{cd}\\
v_{S_{ed}}&=\sqrt{(\hat{x}_e-\hat{x}_d)^2+(\hat{y}_e-\hat{y}_d)^2}-S_{ed}
\end{aligned}$$

$$v_{S_{ae}}=\sqrt{(x_a-\hat{x}_e)^2+(y_a-\hat{y}_e)^2}-S_{ae}$$

$$v_{S_{af}}=\sqrt{(\hat{x}_f-x_a)^2+(\hat{y}_f-y_a)^2}-S_{af}$$

$$v_{S_{fd}}=\sqrt{(\hat{x}_d-\hat{x}_f)^2+(\hat{y}_d-\hat{y}_f)^2}-S_{fd}$$

设测站 A 上的 0 方向为 AB, 测站 B 上的 0 方向为 BC, 测站 C 上的 0 方向为 CD, 测站 D 上的 0 方向为 DE, 测站 E 上的 0 方向为 EA, 测站 F 上的 0 方向为 FD, 则由(6-1-5)式得方向观测值的误差方程为

$$v_{\beta_{af}}=\arctan\frac{\hat{y}_f-y_a}{\hat{x}_f-x_a}-\arctan\frac{y_b-y_a}{x_b-x_a}-\beta_{af}$$

$$v_{\beta_{ae}}=\arctan\frac{\hat{y}_e-y_a}{\hat{x}_e-x_a}-\arctan\frac{y_b-y_a}{x_b-x_a}-\beta_{ae}$$

$$v_{\beta_{ba}}=\arctan\frac{y_a-y_b}{x_a-x_b}-\arctan\frac{\hat{y}_c-y_b}{\hat{x}_c-x_b}-\beta_{ba}$$

$$v_{\beta_{cb}}=\arctan\frac{y_b-\hat{y}_c}{x_b-\hat{x}_c}-\arctan\frac{\hat{y}_d-\hat{y}_c}{\hat{x}_d-\hat{x}_c}-\beta_{cb}$$

$$v_{\beta_{df}}=\arctan\frac{\hat{y}_f-\hat{y}_d}{\hat{x}_f-\hat{x}_d}-\arctan\frac{\hat{y}_e-\hat{y}_d}{\hat{x}_e-\hat{x}_d}-\beta_{df}$$

$$v_{\beta_{dc}}=\arctan\frac{\hat{y}_c-\hat{y}_d}{\hat{x}_c-\hat{x}_d}-\arctan\frac{\hat{y}_e-\hat{y}_d}{\hat{x}_e-\hat{x}_d}-\beta_{dc}$$

$$v_{\beta_{ed}}=\arctan\frac{\hat{y}_d-\hat{y}_e}{\hat{x}_d-\hat{x}_e}-\arctan\frac{y_a-\hat{y}_e}{x_a-\hat{x}_e}-\beta_{ed}$$

$$v_{\beta_{fa}}=\arctan\frac{y_a-\hat{y}_f}{x_a-\hat{x}_f}-\arctan\frac{\hat{y}_d-\hat{y}_f}{\hat{x}_d-x_f}-\beta_{fa}$$

取单位权中误差为 σ_β, 则方向观测值和边长观测值的权分别为

$p_\beta=\frac{\sigma_\beta^2}{\sigma_\beta^2}=1$, $p_{S_i}=\frac{\sigma_\beta^2}{\sigma_{S_i}^2}$. 有了 V 和 P 后, 在 $V'PV=\min$ 下, 用第三章的任一种解算方法, 都能解算出未知数 $\hat{x}_i$ 和 $\hat{y}_i$.

第二节　在工程测量中的应用

施工放样是工程测量中最基本的测量工作. 由于全站仪的普及, 极坐标法放样得到广泛的应用. 极坐标法放样对现场条件要求较低. 只要两个已知点通视, 且在任一已知点上能与待放样点通视即可. 然而, 就这么低的现场条件, 在建筑物密集的地区往往也难以满足. 如图 6-2-1 所示, 两已知点 A、B 彼此通视, 但 A、B 两已知点与待放样点 D 均不通视. 在这种情况下, 可在任意一个同时与 A、B、D 三点通视的地方, 比如图 6-2-2 中的 C 点安置仪器, 通过测量 C 点至 A、B 两点的方向 β_{ca}、β_{cb}和距离 S_{ca}、S_{cb}, 据此计算出放样数据 $\hat{\beta}_{cd}$、$\hat{S}_{cd}$后, 就可放样出 D 点. 这种放样方法通常被称为自由设站法.

在自由设站法放样中, 要通过方向观测值 β_{ca}、β_{cb}和距离观测值 S_{ca}、S_{cb}计算放样数据 $\hat{\beta}_{cd}$、$\hat{S}_{cd}$, 需要先计算出 C 点的坐标 $\hat{x}_c$、$\hat{y}_c$. 为了计算出 C 点的坐标 $\hat{x}_c$、$\hat{y}_c$, 可列如下误差方

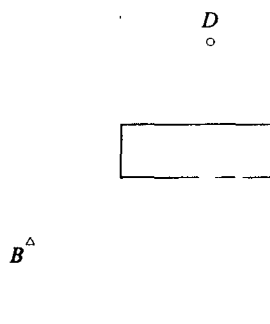


图 6-2-1

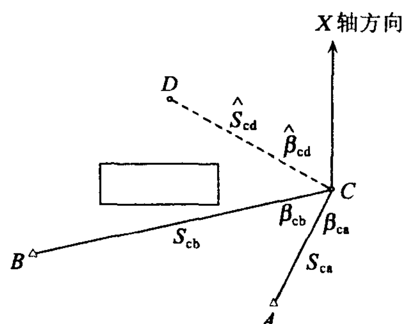


图 6-2-2

程(设 0 方向为 CA 方向)：

$$v_{\beta_{cb}} = \arctan\frac{y_b - \hat{y}_c}{x_b - \hat{x}_c} - \arctan\frac{y_a - \hat{y}_c}{x_a - \hat{x}_c} - \beta_{cb}$$

$$v_{S_{ca}} = \sqrt{(x_a - \hat{x}_c)^2 + (y_a - \hat{y}_c)^2} - S_{ca}$$

$$v_{S_{cb}} = \sqrt{(x_b - \hat{x}_c)^2 + (y_b - \hat{y}_c)^2} - S_{cb}$$

采用第三章中介绍的任一种非线性模型参数估计的解法，均可解算出未知参数 $\hat{x}_c$，$\hat{y}_c$. 在测站 C 上解算出 $\hat{x}_c$，$\hat{y}_c$ 后，由下式即可算出放样数据 $\hat{\beta}_{cd}$，$\hat{S}_{cd}$：

$$\hat{\beta}_{cd} = \arctan\frac{y_d - \hat{y}_c}{x_d - \hat{x}_c}$$

$$\hat{S}_{cd} = \sqrt{(x_d - \hat{x}_c)^2 + (y_d - \hat{y}_c)^2}$$

有了放样数据 $\hat{\beta}_{cd}$，$\hat{S}_{cd}$后，在测站 C 上就可放样出 D 点.

第三节 在摄影测量中的应用

“摄影测量学的主要任务是测制各种比例尺的地形图、建立地形数据库，并为各种地理信息系统和土地信息系统提供基础数据”[72]（李德仁，金为铣，尤兼善，朱宜萱，1995）. 要通过立体像对测制各种比例尺的地形图，首先需要得到被摄物体在地面坐标系中的坐标. 而要得到被摄物体在地面坐标系中的坐标，需要对立体像对进行相对定向和绝对定向. “解析法像对的相对定向是通过计算相对定向元素，建立地面的立体模型”[73]（朱肇光，孙护，崔炳光，1995）. 绝对定向就是把模型点在像空间辅助坐标系中的坐标转化为地面坐标系中的坐标. 要完成这一坐标转换，必须估计出七个坐标转换参数，即摄影测量中的七个绝对定向元素 X_s、Y_s、Z_s、λ、Φ、Ω 和 K. 下面用非线性模型参数估计方法来估计这七个绝对定向元素.

由《摄影测量学》[73]（朱肇光，孙护，崔炳光，1995）知：

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \lambda R \begin{pmatrix} U \\ V \\ W \end{pmatrix} + \begin{pmatrix} X_s \\ Y_s \\ Z_s \end{pmatrix} \tag{6-3-1}$$

式中：X、Y、Z 为控制点在地面坐标系中的已知坐标；U、V、W 为控制点在像空间辅助坐标系中的坐标，是观测值；X_s、Y_s、Z_s 为未知的平移参数；λ 为未知的缩放比例因子；R 为由下式定义的旋转矩阵：

$$R = \begin{pmatrix} \cos\Phi\cos K - \sin\Phi\sin\Omega\sin K & -\cos\Phi\sin K - \sin\Phi\sin\Omega\cos K & -\sin\Phi\cos\Omega \\ \cos\Omega\sin K & \cos\Omega\cos K & -\sin\Omega \\ \sin\Phi\cos K + \cos\Phi\sin\Omega\sin K & -\sin\Phi\sin K + \cos\Phi\sin\Omega\cos K & \cos\Phi\cos\Omega \end{pmatrix} \tag{6-3-2}$$

因为 U、V、W 是观测值，故(6-3-1)式可写为

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \lambda R \begin{pmatrix} U + v_u \\ V + v_v \\ W + v_w \end{pmatrix} + \begin{pmatrix} X_s \\ Y_s \\ Z_s \end{pmatrix} \tag{6-3-3}$$

令

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \lambda R \begin{pmatrix} v_u \\ v_v \\ v_w \end{pmatrix} \tag{6-3-4}$$

则有误差方程

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \lambda R \begin{pmatrix} U \\ V \\ W \end{pmatrix} + \begin{pmatrix} X_s \\ Y_s \\ Z_s \end{pmatrix} - \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{6-3-5}$$

顾及(6-3-2)式知，(6-3-5)式为非线性误差方程．一个控制点可列出 3 个误差方程，当控制点个数大于等于 3 时，误差方程的个数大于等于 9．所以采用第三章介绍的任一种算法，都可以解出这七个绝对定向元素 X_s、Y_s、Z_s、λ、Φ、Ω 和 K．

第四节　在 GPS 定位中的应用

GPS 定位一般分为伪距定位法和载波相位定位法两种．前者的观测值是伪距，后者的观测值是载波相位．需要求解的参数都是测站点 i 的三维坐标 $\hat{x}_i$、$\hat{y}_i$、$\hat{z}_i$．

对于测码伪距绝对定位，其基本方程为

$$\rho_i^j = \sqrt{(x^j - x_i)^2 + (y^j - y_i)^2 + (z^j - z_i)^2} + c\delta t_i + \Delta_i^j \tag{6-4-1}$$

式中：i 为测站点点名；j 为卫星编号；ρ_i^j 为测站点 i 至卫星 j 的伪距；(x^j, y^j, z^j) 为卫星 j 的坐标；(x_i, y_i, z_i) 为待求的测站点 i 的坐标；c 为测距信号的传播速度；$\delta t_i = \Delta t^j - \Delta t_i$ 称为钟差参数；Δ_i^j 为包含电离层误差、大气层误差、随机误差等的真误差．将(6-4-1)式写成误差方程的形式为

$$v_i^j = \sqrt{(x^j - \hat{x}_i)^2 + (y^j - \hat{y}_i)^2 + (z^j - \hat{z}_i)^2} + c\delta\hat{t}_i - \rho_i^j \tag{6-4-2}$$

式中有 3 个坐标参数 $(\hat{x}_i, \hat{y}_i, \hat{z}_i)$ 和 1 个钟差参数 $\delta\hat{t}_i$．当同时接收到 4 颗以上卫星的信号时，用第三章中介绍的任一种解算方法，都能解算出未知数 $(\hat{x}_i, \hat{y}_i, \hat{z}_i)$．

对于测相伪距绝对定位，其基本方程为

$$\varphi_i^j = \frac{f}{c}\sqrt{(x^j - x_i)^2 + (y^j - y_i)^2 + (z^j - z_i)^2} + f\delta t_i - N_i^j + \Delta_i^j \tag{6-4-3}$$

式中：φ_i^j 为载波相位观测值；f 为信号的频率；N_i^j 为整周未知数；其余符号的意义同(6-4-1)式.

与(6-4-3)式相应的误差方程为

$$v_i^j = \frac{f}{c}\sqrt{(x^j - \hat{x}_i)^2 + (y^j - \hat{y}_i)^2 + (z^j - \hat{z}_i)^2} + f\delta\hat{t}_i - N_i^j - \varphi_i^j \tag{6-4-4}$$

式中有 3 个坐标参数($\hat{x}_i, \hat{y}_i, \hat{z}_i$)、1 个钟差参数 $\delta\hat{t}_i$ 和一个整周模糊度参数. 当观测几个历元后, 整周模糊度参数可以通过专门的方法固定下来. 因此, 只要同时接收到 4 颗以上卫星的信号时, 用第三章中介绍的任一种解算方法, 都能解算出未知数($\hat{x}_i, \hat{y}_i, \hat{z}_i$).

在 GPS 测量中, 用得最多的是载波相位相对定位. 在 GPS 相对定位中, 常常对观测值进行线性组合. 当前普遍应用的重要组合形式, 只有单差、双差和三差三种[74](周忠谟、易杰军, 周琪, 1997). 在此只讨论单差的情况. 所谓单差, 就是在两测站之间对同步观测值求差. 其观测方程为

$$\Delta\varphi^j = \frac{f}{c}\left(\sqrt{(x^j - x_2)^2 + (y^j - y_2)^2 + (z^j - z_2)^2} - \sqrt{(x^j - x_1)^2 + (y^j - y_1)^2 + (z^j - z_1)^2}\right) + f\Delta t - \Delta N^j + \Delta^j$$

式中：$\Delta\varphi^j = \varphi_2^j - \varphi_1^j$, $\Delta t = \delta t_2 - \delta t_1$, $\Delta N^j = N_2^j - N_1^j$.

由于相对定位中有一个测站为已知点, 比如测站 1 为已知点, 则测站 1 至卫星 j 的距离已知, 即 $\rho_1^j = \sqrt{(x^j - x_1)^2 + (y^j - y_1)^2 + (z^j - z_1)^2}$ = 常数. 于是单差法的误差方程为

$$v^j = \frac{f}{c}\sqrt{(x^j - \hat{x}_2)^2 + (y^j - \hat{y}_2)^2 + (z^j - \hat{z}_2)^2} + f\delta\hat{t}_i - N_i^j - \left(\Delta\varphi^j + \frac{f}{c}\rho_1^j\right)$$

当同时接收到 4 颗以上卫星的信号时, 用第三章中介绍的任一种解算方法, 都能解算出未知数($\hat{x}_2, \hat{y}_2, \hat{z}_2$).

附录 A　立体阵的定义、运算及其性质

我们知道,m 维向量函数$f(X)$关于 n 维向量X 的一阶导数是一个$m\times n$ 阶的矩阵函数.那么 $f(X)$关于 X 的二阶偏导数是什么呢? 由于矩阵函数$\frac{\partial f}{\partial X}$的任一行都是一个$n$ 维向量$\frac{\partial f_{ij}(X)}{\partial X}(j=1,2,\cdots,n)$这个 n 维向量函数$\frac{\partial f_{ij}(X)}{\partial X}$关于$X$ 的一阶偏导数又是$n\times n$ 阶的矩阵.因此,m 个n 维向量关于X 的一阶偏导数共有m 个$n\times n$ 阶矩阵.若将这 m 个$n\times n$ 阶矩阵按顺序上下叠置,就得到一个 m 层的三维数组,这样的三维数组称为立体阵(Array).

立体阵及其方括号乘法,最早于 1980 年由 Bates 和 Watts 在他们的论文中提出[75](Bates & Watts, 1980).1983 年 Tsai 在他的博士论文中对其进行过初步整理[76](Tsai, 1983).1986 年我国学者韦博成在 Tsai 的基础上进行了系统的总结和扩充[77](韦博成, 1986).立体阵在非线性模型参数估计中占有非常重要的地位,现介绍如下.

一、关于立体阵的若干定义

定义 A-1　称 $n\times p\times q$ 的三维数组$X=(X_{kij})$为立体阵,简称立阵.其中 n 表示此立阵共有 n 层,$p\times q$ 表示每层都是一个$p\times q$ 阶的矩阵.X_{kij}为立阵X 的一个元素.下标 k、i、j 分别表示该元素所在的层、行和列.即 X_{kij}表示立阵X 的第k 层、第 i 行、第 j 列上的元素,见图 A-1.

可采用下列两种方式来理解立体阵:

第一种方式,立体阵 X 表示一个元素为向量的矩阵,该矩阵中的任一元素为 $X_{ij}=(x_{1ij}\quad x_{2ij}\quad\cdots\quad x_{nij})'$.

第二种方式,立体阵 X 由n 个$p\times q$ 阶的矩阵$X_k(k=1,2,\cdots,n)$上下叠置而成.X_k 表示立体阵X 中第k 层的那个矩阵.例如,一个 $3\times2\times2$ 的立阵为

$$X=\left\{\begin{array}{c}\begin{array}{ccc} & 1 \text{——} 3 \\ / & & / \\ 5 \text{——} 7 & \end{array}\\[6pt] \begin{array}{ccc} & 2 \text{——} 4 \\ / & & / \\ 6 \text{——} 8 & \end{array}\\[6pt] \begin{array}{ccc} & 9 \text{——} 11 \\ / & & / \\ 13 \text{——} 15 & \end{array}\end{array}\right\}$$

简写为

$$X_1=\begin{pmatrix}1&3\\5&7\end{pmatrix},X_2=\begin{pmatrix}2&4\\6&8\end{pmatrix},X_3=\begin{pmatrix}9&11\\13&15\end{pmatrix}$$

而 $X_{212}=4$

定义 A-2　若两个立阵 X 和 Y 的对应元素一一相等,则这两个立阵相等,即当 $X_{kij}=Y_{kij}$时,有 $X=Y$.

定义 A-3　设 X 和 Y 均为$n\times p\times q$ 阶立阵,定义 X 与 Y 之和(差)为

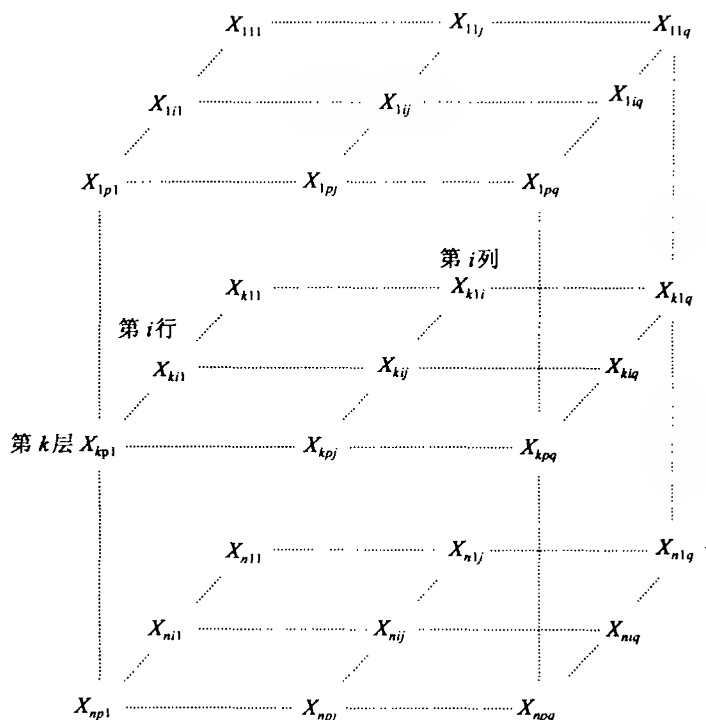


图 A-1

$$Z = X \pm Y$$

各元素的关系为

$$Z_{kij} = X_{kij} \pm Y_{kij} \tag{A-1}$$

定义 A-4 设 A 为 $r \times p$ 阶矩阵, X 为 $n \times p \times q$ 阶立阵. 矩阵 A 左乘立阵 X, 等于 A 左乘 X 的每一层的矩阵, 然后将 n 个乘积矩阵按 X 的相同顺序叠放. 其维数关系为

$$Y = AX \tag{A-2}$$

Y 中各元素的计算公式为

$$Y_{kij} = \sum_{s=1}^{p} a_{is} x_{ksj} \tag{A-3}$$

注意: 只有当 A 的列数(A 的第二个下标)与 X 的行数(X 的第二个下标)相同时, A 与 X 才可乘. 仿此可类似地定义矩阵与立阵的右乘.

例 A-1 已知

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix},\quad X = \left[\begin{array}{c} \begin{matrix} & 2 & 1 \\ 1 & 2 & \end{matrix} \\ \\ \begin{matrix} & 3 & 4 \\ 4 & 3 & \end{matrix} \\ \\ \begin{matrix} & 6 & 5 \\ 5 & 6 & \end{matrix} \end{array}\right]$$

求 A 左乘X 的积 Y.

$$Y = AX = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} \begin{array}{cccc} & 2 & & 1 \\ 1 & & 2 & \\ & 3 & & 4 \\ 4 & & 3 & \\ & 6 & & 5 \\ 5 & & 6 & \end{array} \end{pmatrix} = \begin{pmatrix} \begin{array}{cccc} & 4 & & 5 \\ 5 & & 4 & \\ & 11 & & 10 \\ 10 & & 11 & \\ & 16 & & 17 \\ 17 & & 16 & \end{array} \end{pmatrix}$$

当 A 为 $1\times p$ 维行向量, X 为 $n\times p\times q$ 阶立阵时, A 左乘X 得到一个 $n\times q$ 的矩阵.

例 A-2 已知 $A=(1\quad 2)$, X 如例 2-2-1 中所示, 求 A 左乘X 的积 Y.

$$Y = AX = (1\quad 2) \begin{pmatrix} \begin{array}{cccc} & 2 & & 1 \\ 1 & & 2 & \\ & 3 & & 4 \\ 4 & & 3 & \\ & 6 & & 5 \\ 5 & & 6 & \end{array} \end{pmatrix} = \begin{pmatrix} 4 & 5 \\ 11 & 10 \\ 16 & 17 \end{pmatrix}$$

特别地, 当 A 为 $p\times 1$ 维列向量, X 为 $n\times p\times p$ 阶立阵时, 二次型 $A'XA$ 表示一个 n 维列向量, 即

$$\begin{aligned} A'XA &= (A'X_1A \quad A'X_2A \quad \cdots \quad A'X_nA)' \\ &= \left(\sum_{i=1}^{p}\sum_{j=1}^{p} x_{1ij}a_ia_j \quad \sum_{i=1}^{p}\sum_{j=1}^{p} x_{2ij}a_ia_j \quad \cdots \quad \sum_{i=1}^{p}\sum_{j=1}^{p} x_{nij}a_ia_j \right)' \end{aligned} \tag{A-4}$$

例 A-3 已知 $A=(1\quad 2)'$, X 如例 2-2-1 中所示, 求二次型 $A'XA$.

$$\begin{aligned} Y = A'XA &= (1\quad 2) \begin{pmatrix} \begin{array}{cccc} & 2 & & 1 \\ 1 & & 2 & \\ & 3 & & 4 \\ 4 & & 3 & \\ & 6 & & 5 \\ 5 & & 6 & \end{array} \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \\ &= (14\quad 31\quad 50)' \end{aligned}$$

定义 A-5 设 X 为 $n\times p\times p$ 阶立阵, X 的迹定义为一个 n 维列向量, 记为 $\mathrm{tr}(X)$. 则

$$\mathrm{tr}(X) = (\mathrm{tr}(X_1) \quad \mathrm{tr}(X_2) \quad \cdots \quad \mathrm{tr}(X_n))' \tag{A-5}$$

例 A-4 已知 $X = \begin{pmatrix} \begin{array}{cccc} & 8 & & 5 \\ 7 & & 9 & \\ & 6 & & 8 \\ 3 & & 2 & \end{array} \end{pmatrix}$, 求 X 的迹.

$$\mathrm{tr}(X) = (\mathrm{tr}(X_1) \quad \mathrm{tr}(X_2))' = (17\quad 8)' \tag{A-6}$$

定义 A-6 设 X 为 $n\times p\times p$ 阶立阵, X 的向量表示定义为一个 $pq\times n$ 阶矩阵, 记为 $\mathrm{Vec}(X)$. 该矩阵在$((j-1)p+i, k)$处的元素是 x_{kij}.

例 A-5 已知立阵 X 如例 A-1，求 X 的向量表示 $\mathrm{Vec}(X)$.

$$\mathrm{Vec}(X)=Y=\begin{pmatrix}2&3&6\\1&4&5\\1&4&5\\2&3&6\end{pmatrix}$$

当 $j=2, i=1, k=2$ 时，$y_{p(j-1)+i,k}=y_{32}=x_{212}=4$.

定义 A-7 设 A 为 $m\times n$ 阶矩阵，X 为 $n\times p\times q$ 阶立阵. A 与 X 的方括号乘积 $Y=[A][X]$ 表示矩阵 A 与立阵 X 的层的乘法，即

$$y_{sij}=\sum_{k=1}^{n}a_{sk}x_{kij} \tag{A-7}$$

其维数关系为

$$\underset{m\times p\times q}{Y}=[\underset{m\times n}{A}][\underset{n\times p\times q}{X}] \tag{A-8}$$

仿此类似地定义方括号右乘.

例 A-6 已知 $\underset{4\times 3}{A}=\begin{pmatrix}1&2&3\\3&4&5\\5&6&7\\7&8&9\end{pmatrix}$，$\underset{3\times 2\times 2}{X}$ 如例 A-1 求 A 与 X 的方括号乘积 Y.

$$y=[A][x]=\left[\begin{pmatrix}1&2&3\\3&4&5\\5&6&7\\7&8&9\end{pmatrix}\right]\left[\begin{pmatrix}\begin{matrix}2&—&1\\1&—&2\end{matrix}\\ \\ \begin{matrix}4&—&3\\3&—&4\end{matrix}\\ \\ \begin{matrix}6&—&5\\5&—&6\end{matrix}\end{pmatrix}\right]=\begin{pmatrix}\begin{matrix}28&—&22\\22&—&28\end{matrix}\\ \\ \begin{matrix}52&—&40\\40&—&52\end{matrix}\\ \\ \begin{matrix}76&—&58\\58&—&76\end{matrix}\\ \\ \begin{matrix}100&—&76\\76&—&100\end{matrix}\end{pmatrix}$$

定义 A-8 设 $X=(x_{kij})$ 为 $n\times p\times q$ 阶立阵，则 X 的转置定义为 $X'=(x_{kji})$ 为 $n\times q\times p$ 阶立阵.

例 A-7 已知 $X=\begin{pmatrix}\begin{matrix}3&\cdots&5&\cdots&7\\2&\cdots&4&\cdots&6\end{matrix}\\ \\ \begin{matrix}1&\cdots&8&\cdots&9\\11&\cdots&13&\cdots&15\end{matrix}\end{pmatrix}$，求 X'.

$$X' = \begin{pmatrix} & 3 & \cdots & 2 \\ 5 & \cdots & 4 & \\ 7 & \cdots & 6 & \\ & 1 & \cdots & 11 \\ 8 & \cdots & 13 & \\ 9 & \cdots & 15 & \end{pmatrix}$$

二、各种运算的基本性质

以上所定义的立体阵的各种运算具有如下基本性质(X、Y、Z 为立阵):

1. $[I][X]=X$
2. $[\lambda A][X]=[A][\lambda X]=\lambda[A][X]$ (λ 为实数)
3. $[A+B][X]=[A][X]+[B][X]$
4. $[A][X+Y]=[A][X]+[A][Y]$
5. $[A][X]=[X][A']$
6. $[A][LXM]=L[A][X]M$

证明: 设 $\underset{n\times p\times q}{Y}=\underset{p\times p}{L}\underset{n\times p\times q}{X}$, $\underset{n\times p\times q}{Z}=[\underset{m\times p}{A}][\underset{m\times p}{L}\underset{n\times p\times q}{X}]=[A][Y]$

$$\underset{m\times p\times q}{W}=[A][X],\quad \underset{n\times p\times q}{U}=L[A][X]=LW$$

由(A-3)式知: $y_{kij}=\sum_{e=1}^{p} l_{ie}x_{kej}$

由(A-7)式知: $z_{sij}=\sum_{k=1}^{n} a_{sk}y_{kij}$

$$=\sum_{k=1}^{n}\sum_{e=1}^{p} a_{sk}l_{ie}x_{kej}$$

$$w_{sij}=\sum_{e=1}^{p} a_{sk}l_{ie}x_{kej}$$

而 $u_{sij}=\sum_{e=1}^{p} l_{ie}w_{kej}=\sum_{e=1}^{p} l_{ie}\sum_{s=1}^{n} a_{sk}l_{iej}=z_{sij}$

所以

$$[A][LX]=L[A][X] \tag{A-9}$$

同理, 当 M 为 $q\times q$ 阶方阵时, 有

$$[A][XM]=[A][X]M \tag{A-10}$$

综合(A-9)式和(A-10)式得

$$[A][LXM]=L[A][X]M$$

性质 6 得证.

7. $[AB][X]=[A][[B][X]]$

证明: 设

$$\underset{m\times n}{C}=\underset{m\times r}{A}\underset{r\times n}{B},\ \underset{m\times p\times q}{Y}=[\underset{m\times n}{C}][\underset{n\times p\times q}{X}],\ \underset{r\times p\times q}{Z}=[\underset{r\times n}{B}][\underset{n\times p\times q}{X}],\ \underset{m\times p\times q}{W}=[\underset{m\times r}{A}][\underset{n\times p\times q}{Z}]$$

由(A-7)式知

$$y_{sij} = \sum_{e=1}^{p} c_{sk} x_{kij} = \sum_{k=1}^{n} \sum_{e=1}^{r} a_{se} b_{ek} x_{kij}$$

$$z_{eij} = \sum_{k=1}^{n} b_{ek} x_{kij}$$

$$w_{sij} = \sum_{e=1}^{r} a_{se} z_{eij} = \sum_{e=1}^{r} \sum_{k=1}^{n} a_{se} b_{ek} x_{kij} = y_{sij}$$

即 $Y = W$. 所以性质 7 成立.

8. $[Aa][(Bb)(Cc)'] = B[A][[a][bc']]C'$

9. $d'Xd = (Xd)d = \sum_{i=1} \sum_{j=1} d_i d_j x_{ij}$

其中: d 为 $p \times 1$ 维行向量, X 为 $n \times p \times p$ 阶立阵.

证明: 令

$$\underset{n\times p\times 1}{Y} = \underset{n\times p\times p}{X}\ \underset{p\times 1}{d},\quad \underset{n\times 1\times 1}{W} = (Xd)d = Yd,\quad \underset{n\times 1\times 1}{Z} = d'Y$$

$$y_{sj} = \sum_{k=1}^{p} x_{sjk} d_{kj},\ W_s = \sum_{j=1}^{p} y_{sj} d_j = \sum_{j=1}^{p} \sum_{k=1}^{p} x_{sjk} d_k d_j$$

而 $Z_s = \sum_{j=1}^{p} d_j y_{sj} = \sum_{j=1}^{p} d_j \sum_{k=1}^{p} x_{sjk} d_k = \sum_{j=1}^{p} \sum_{k=1}^{p} x_{sjk} d_k d_j = W_s$

所以性质 9 成立.

10. $A(d'Xd) = [A][d'Xd] = d'[A][X]d$

其中: d 为 $p \times 1$ 维行向量.

11. $\mathrm{tr}(AX) = \mathrm{tr}(XA)$

12. $A\,\mathrm{tr}(X) = \mathrm{tr}([A][X])$

证明: 令 $\underset{n\times n\times n}{Y} = [\underset{n\times n}{A}][\underset{n\times n\times n}{X}]$

$$\underset{n\times 1}{b} = \mathrm{tr}([A][x]) = \mathrm{tr}(Y),\ \underset{n\times 1}{a} = A\,\mathrm{tr}(X) = AC,\ \underset{n\times 1}{C} = \mathrm{tr}(X)$$

由(A-7)式知: $y_{sii} = \sum_{k=1}^{n} a_{sk} x_{kii}$, 而 $C_k = \sum_{i=1}^{n} y_{sii} = \sum_{i=1}^{n} x_{kii}$

$$b_s = \mathrm{tr}(y_s) = \sum_{i=1}^{n} y_{sii} = \sum_{i=1}^{n} \sum_{k=1}^{n} a_{sk} x_{kii}$$

则

$$a_s = \sum_{k=1}^{n} a_{sk} c_k = \sum_{k=1}^{n} a_{sk} \sum_{i=1}^{n} x_{kii} = \sum_{i=1}^{n} \sum_{k=1}^{n} a_{sk} x_{kii} = b_s$$

所以性质 12 成立.

13. $\mathrm{Vec}([A][X]) = (\mathrm{Vec}(X))A'$

证明: 设 A 为 $m \times n$ 阶矩阵, X 为 $n \times p \times q$ 阶立阵. $Y = [A][X]$, 由(A-7)式知

$$y_{kij} = \sum_{s=1}^{n} a_{ks} x_{sij}$$

则由定义 A-6 知 $\mathrm{Vec}([A][X])$ 在 $((j-1)p+i, k)$ 处的元素为:

$$y_{kij} = \sum_{s=1}^{n} a_{ks} x_{sij}$$

由定义 A-6 知

$$\operatorname{Vec}(X)=\begin{pmatrix} x_{111} & x_{211} & \cdots & x_{n11} \\ x_{112} & x_{212} & \cdots & x_{n12} \\ \vdots & \vdots & & \vdots \\ x_{1pq} & x_{2pq} & \cdots & x_{npq} \end{pmatrix}$$

$$(\operatorname{Vec}(X))A'=\begin{pmatrix} x_{111} & x_{211} & \cdots & x_{n11} \\ x_{112} & x_{212} & \cdots & x_{n12} \\ \vdots & \vdots & & \vdots \\ x_{1pq} & x_{2pq} & \cdots & x_{npq} \end{pmatrix}\begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}$$

$$=\begin{pmatrix} \sum_{s=1}^{n} x_{s11}a_{1s} & \sum_{s=1}^{n} x_{s11}a_{2s} & \cdots & \sum_{s=1}^{n} x_{s11}a_{ms} \\ \sum_{s=1}^{n} x_{s12}a_{1s} & \sum_{s=1}^{n} x_{s12}a_{2s} & \cdots & \sum_{s=1}^{n} x_{s12}a_{ms} \\ \vdots & \vdots & & \vdots \\ \sum_{s=1}^{n} x_{spq}a_{1s} & \sum_{s=1}^{n} x_{spq}a_{2s} & \cdots & \sum_{s=1}^{n} x_{spq}a_{ms} \end{pmatrix}$$

显然在$((j-1)p+i,k)$处的元素为 $\sum_{s=1}^{n} x_{sij}a_{ks}$，所以性质 13 成立.

14. $\operatorname{Vec}(AXB)=(B'\otimes A)\operatorname{Vec}(X)$

15. $([A][X])'=[X'][A']$

以上仅就本书所用到的立体阵运算作了一些一般定义，并总结了 15 条性质. 由于立体阵的运算变化很多，因此还可定义很多种运算，并总结各自的性质. 读者可根据实际需要自己定义和总结.

附录B　蒙特卡罗积分

在贝叶斯估计中，往往要进行非常复杂的多重积分计算．由于被积函数非常复杂，无法用解析方法求出积分，故一般都要借助数值积分法．尽管存在很多数值积分方法，但这些方法计算多重积分，工作量大，速度缓慢，效率很低，而蒙特卡罗积分却能帮助我们克服这些困难．因此，在贝叶斯估计中常常使用蒙特卡罗积分．蒙特卡罗积分是蒙特卡罗方法引入计算数学的开端．在实际中，许多需要计算多重积分的复杂问题，用蒙特卡罗方法一般都能很有效地予以解决．为此，本节扼要介绍蒙特卡罗方法．

一、蒙特卡罗方法的基本思想

蒙特卡罗方法(Monte Carlo Method)的定名和系统发展约始于20世纪40年代中期．但如果从方法特征的角度来说(尽管在当时方法雏形的出现是孤立的，而且也没有得到发展)，可以一直追溯到19世纪后半叶的蒲丰(Buffon)随机投针试验，即著名的所谓蒲丰问题．

蒙特卡罗方法又称随机模拟(Random Simulation)方法，有时也称随机抽样(Random Sampling)技术或统计试验(Statistical Testing)方法．它的基本思想是，为了求解数学、物理、工程技术以及生产管理等方面的问题，首先建立一个概率模型或随机过程，使它的参数等于问题的解．然后通过模型或过程的观察或抽样试验来计算所求参数的统计特征，最后给出所求解的近似值．而解的精确度可用估计值的标准误差来表示．

用蒙特卡罗方法求解时，最简单的情况是模拟一个发生概率为 p 的随机事件 A．考虑一个随机变量 ξ，若在一次试验中事件 A 出现，则 ξ 取值为1，若事件 A 不出现，则 ξ 取值为0．令 $q=1-p$，那么随机变量 ξ 的数学期望为 $\mathrm{E}(\xi)=1\cdot p+0\cdot q=p$，此即一次试验中事件 A 出现的概率．ξ 的方差为 $\mathrm{E}(\xi-\mathrm{E}(\xi))^2=p-p^2=pq$．假设在 N 次试验中，事件 A 出现 v 次，那么观察频数 v 也是一个随机变量，其数学期望 $\mathrm{E}(v)=Np$，方差 $\sigma_v^2=Npq$，令 $\bar{p}=\frac{v}{N}$，表示观察频率，那么按照强大数定理，当 N 充分大时，下式

$$\bar{p}=\frac{v}{N}\approx\mathrm{E}(\xi)=p \tag{B-1}$$

成立的概率为1．因此，由上述模型得到的频率 $\bar{p}=\frac{v}{N}$ 近似地等于所求量 p．这就说明了频率收敛于概率．而且可用样本方差．

$$\sigma_p^2=\frac{\bar{p}(1-\bar{p})}{N-1} \tag{B-2}$$

作为理论方差 σ_p^2 的估值．

蒙特卡罗方法可以解决各种类型的问题，但总的来说，视其是否涉及随机过程的性态和

结果.用蒙特卡罗方法处理的问题可以分为两类:

第一类是确定性的数学问题.用蒙特卡罗方法求解这类问题的方法是,首先建立一个与所求解有关的概率模型,使所求的解就是我们所建立模型的概率分布或数学期望.然后对这个模型进行随机抽样观察,即产生随机变量;最后用其算术平均值作为所求解的近似估计值.计算多重积分、矩阵求逆、解线性方程组等都属于这一类.

第二类是随机性问题.例如电子在介质中的扩散等问题,就属于随机性问题.对于这类问题,虽然有时可表示为多重积分或某些函数方程,进而考虑用随机抽样方法求解,然而一般情况下都不采用这种间接模拟方法,而是采用直接模拟方法,即根据实际物理情况的概率法则,用电子计算机进行抽样试验.原子核物理问题、运筹学中的库存问题,随机服务系统中的排队问题等都属于这一类.

二、蒙特卡罗方法的特点

蒙特卡罗方法是一种具有独特风格的数值计算方法.其优点以及与其他数值计算方法的不同点,可归纳为以下几个方面.

(1)蒙特卡罗方法及其程序结构简单

例如,用蒙特卡罗方法计算积分,只需做大量简单的重复抽样,而抽样的方法和程序都是很简单的.

又例如用随机游动方法求解椭圆型差分方程边值问题时,可以只求解我们所需要的某个点上的值,而不需要求出全部网络点上的值.

(2)收敛的概率性和收敛速率与问题维数无关

蒙特卡罗方法的收敛是概率意义下的收敛.其收敛速率与一般数值方法相比是很慢的,故蒙特卡罗方法不宜解决精度要求很高的问题.

蒙特卡罗方法的误差 ε 只与标准差 σ 和样本容量 N 有关,而与样本中元素所在空间无关,即蒙特卡罗方法的收敛速度与问题的维数无关.其他数值方法则不然,这就决定了蒙特卡罗方法对多维问题的适用性.

(3)蒙特卡罗方法的适应性强

蒙特卡罗方法广泛的适应性是不可忽视的,而且是很重要的,应用蒙特卡罗方法解题,受问题条件限制的影响较小.

三、蒙特卡罗积分

用蒙特卡罗方法计算定积分,有着十分重要的意义,因为在实际工作中近似地计算定积分是经常碰到的数学问题,对于复杂的单重与多重定积分,人们通常选用矩形公式、辛普森公式等来完成积分的近似计算.在许多情况下,使用这些近似公式虽然能得到相当满意的结果,但计算量随着积分重数的增加而显著增加,以致达到电子计算机都难以完成的程度.使用蒙特卡罗方法计算积分,将不存在这个问题,而蒙特卡罗方法积分的误差与积分重数无关.

设有一 S 重积分

$$I=\int\cdots\int_{\Omega} g(x_1,x_2,\cdots,x_s)\mathrm{d}x_1\cdots\mathrm{d}x_s \tag{B-3}$$

其中 Ω 为S 维积分域,考虑 Ω 上的一个概率密度函数$f(x_1,x_2,\cdots,x_s)$,它满足如下条件:

$$f(x_1,x_2,\cdots,x_s)\neq 0$$

当$(x_1,x_2,\cdots,x_s)\in\Omega$, $g(x_1,x_2,\cdots,x_s)\neq 0$ 时,令

$$g^*(x_1,x_2,\cdots,x_s)=\begin{cases}\dfrac{g(x_1,x_2,\cdots,x_s)}{f(x_1,x_2,\cdots,x_s)} & f(x_1,x_2,\cdots,x_s)\neq 0\\ 0 & f(x_1,x_2,\cdots,x_s)=0\end{cases} \tag{B-4}$$

于是(B-3)式可改写为

$$\begin{aligned}I&=\int\cdots\int_{\Omega} g^*(x_1,x_2,\cdots,x_s)f(x_1,x_2,\cdots,x_s)\mathrm{d}x_1,\mathrm{d}x_2,\cdots,\mathrm{d}x_s\\&=\mathrm{E}[g^*(x_1,x_2,\cdots,x_s)]\end{aligned} \tag{B-5}$$

(B-5)式表明,S 重积分I 是随机变量函数$g^*(x_1,x_2,\cdots,x_s)$的数学期望.如果抽选服从$f(x_1,x_2,\cdots,x_s)$的 N 个点$(x_{i1},x_{i2},\cdots,x_{is})$, $i=1,2,\cdots,N$ 并构成N 个函数值$g^*(x_{i1},x_{i2},\cdots,x_{is})$,那么就可用其算术平均值

$$\bar{I}=\frac{1}{N}\sum_{i=1}^{N} g^*(x_{i1},x_{i2},\cdots,x_{is}) \tag{B-6}$$

作为 I 的近似值,即

$$I\approx\bar{I}=\frac{1}{N}\sum_{i=1}^{N}\frac{g(x_{i1},x_{i2},\cdots,x_{is})}{f(x_{i1},x_{i2},\cdots,x_{is})} \tag{B-7}$$

(B-7)式称为蒙特卡罗积分.由(B-7)式知,计算蒙特卡罗积分的主要问题是要寻找一个合适的密度函数 $f(x_1,x_2,\cdots,x_s)$.而且 $f(x_1,x_2,\cdots,x_s)$应很接近原被积函数 $g(x_1,x_2,\cdots,x_s)$,使其 $g^*(x_1,x_2,\cdots,x_s)$几乎为一常数.显然,这一要求是很难满足的.解决这个问题的最简单的方法是在积分域 Ω 上均匀地抽选N 个数据点$(x_{i1},x_{i2},\cdots,x_{is})$.这就意味着 $f(x_{i1},x_{i2},\cdots,x_{is})$是 Ω 上的均匀分布的密度函数.于是有

$$f(x_1,x_2,\cdots,x_s)=\begin{cases}\dfrac{1}{V_s} & (x_1,x_2,\cdots,x_s)\in\Omega\\ 0 & 其他\end{cases} \tag{B-8}$$

式中:V_s 代表积分区域的体积.

将(B-8)式代入(B-7)式,得

$$I\approx\bar{I}=\frac{V_s}{N}\sum_{i=1}^{N} g(x_{i1},x_{i2},\cdots,x_{is}) \tag{B-9}$$

特别地,当积分域 Ω 平行于坐标轴时,均匀分布的密度函数(B-8)式可写为

$$f(x_1,x_2,\cdots,x_s)=\begin{cases}\displaystyle\prod_{j=1}^{s}\frac{1}{b_j-a_j} & a_j\leqslant x_j\leqslant b_j\\ 0 & x_j<a_j,x_j<b_j\end{cases} \tag{B-10}$$

于是(B-9)式变为

$$I\approx\bar{I}=\Big[\prod_{j=1}^{s}(b_j-a_j)\Big]\frac{1}{N}\sum_{i=1}^{N} g(x_{i1},x_{i2},\cdots,x_{is}) \tag{B-11}$$

由蒙特卡罗积分方法[78](徐钟济,1985)知,用蒙特卡罗方法对(B-9)式或(B-11)式计

算定积分时,其误差的阶为$0(N^{-\frac{1}{2}})$,它与积分重数无关.而用矩形公式求 S 重定积分时,误差的阶为$0(N^{-\frac{1}{S}})$.这里 N 表示求积分结点数.因此,当 $S>3$ 时,使用蒙特卡罗方法求积分就开始显示出其优越性了.

附录C 解非线性方程组的一类离散的Newton算法

(陈志,高旅端,邓乃扬)

一、引言

考虑非线性方程组

$$F(x)=0,\quad F(x)=(f_1(x)\quad f_2(x)\quad \cdots\quad f_n(x))^{\mathrm{T}} \tag{C-1-1}$$

设 x_i 是当前的迭代点,为计算下一个迭代点,Newton 法是求解方程

$$F(x_i)+F'(x_i)(x-x_i)=0 \tag{C-1-2}$$

若用差商代替导数,离散 Newton 法要解如下方程

$$F(x_i)+J(x_i,h)(x-x_i)=0 \tag{C-1-3}$$

其中

$$J(x_i,h)=\frac{1}{h}(f_i(x_i+he_k)-f_i(x_i))$$

这里为了计算 $J(x_i,h)$,需要计算 n^2 个函数值.为了提高效能,Brown 方法使用代入消元的办法来减少函数值计算量.它是再通过一次内迭代从 x_i 得到下一个迭代点 x_{i+1}.

设 $x_i=(\xi_1\quad \xi_2\quad \cdots\quad \xi_n)^{\mathrm{T}}$, $t=(t_1\quad t_2\quad \cdots\quad t_n)^{\mathrm{T}}$, t 为变量.Brown 方法的基本思想如下:

对 $f_1(x_i)$在 x_i 处作线性近似

$$f_1(x_i)+f_1'(x_i)^{\mathrm{T}}(t-x_i)=0 \tag{C-1-4}$$

解出

$$t_1-\xi_1=L_1(t_2-\xi_2,\cdots,t_n-\xi_n) \tag{C-1-5}$$

然后代入第二个函数,得到

$$f_2(L_1+\xi_1,t_2,\cdots,t_n)=g_2(t_2,\cdots,t_n) \tag{C-1-6}$$

这是关于 $t_2,\cdots,t_n$ 的函数.当$(t_2\quad t_3\quad \cdots\quad t_n)^{\mathrm{T}}=(\xi_2\quad \xi_3\quad \cdots\quad \xi_n)^{\mathrm{T}}$ 时,由(C-1-4)、(C-1-5)式得到

$$t_1=L_1(0,\cdots,0)+\xi_1=-\frac{f_1(x_i)}{f_{1,1}'(x_i)}+\xi_1 \tag{C-1-7}$$

---

* 本附录系原文摘录,故矩阵 A 的转置用 A^{T} 表示;F 的导数用 F′表示。与本书的符号不同,特此说明。

这里 $f_{1,1}'$表示f_1 对第一个分量的偏导数.由(C-1-6)式得到

$$g_2(\xi_2,\xi_3,\cdots,\xi_n)=f_2\left(-\frac{f_1(x_i)}{f_{1,1}'(x_i)}+\xi_1,\xi_2,\cdots,\xi_n\right) \tag{C-1-8}$$

显然

$$g_2(\xi_2,\xi_3,\cdots,\xi_n)\neq f_2(\xi_1,\xi_2,\cdots,\xi_n) \tag{C-1-9}$$

同样当在$(\xi_2\quad \xi_3\quad \cdots\quad \xi_n)^{\mathrm{T}}$ 处展开,可解出

$$t_2-\xi_2=L_2(t_3-\xi_3,\cdots,t_n-\xi_n)$$

以此类推,通过回代解得 $t=x_{i+1}$.

当用差商计算导数及考虑主元选取时,上述方法是离散 Brown 方法.完成从 x_i 到x_{i+1} 的一次主迭代时,函数值的计算量得以减少,约为离散 Brown 法的一半.若利用坐标旋转,选取合适的坐标系,能够得到离散 Brown 方法.

Brown 方法和 Brent 方法属于二次迭代 Newton 型算法.当差商的步长 $h\to0$ 时,Brown 方法和 Brent 方法实际求解的方程并不是对应的在 x_i 处的 Newton 方程,而是如下线性方程组:

$$f_j(y_j)+f_j'(y_j)^{\mathrm{T}}(y-y_j)=0\quad(j=1,2,\cdots,n) \tag{C-1-10}$$

这里 $y_j=x_i$.(C-1-6)、(C-1-9)式也能表明这一点.

当$\|F(x_i)\|$值较大时,y_j 之间的距离也较大,从(C-1-8)式看出约与$\dfrac{\|F(x_i)\|}{\|F'(x_i)\|}$同阶,因此它们收敛更依赖于初值.第三节中的数值例子表明了这点.数值实验表明,当维数增加时,总体看收敛情况比离散 Newton 法要差.

我们希望直接从 Newton 方程(C-1-2)式获得一种算法,当用差商代替导数时,其效能与 Brown 方法相同,而当步长 $h\to0$ 时,求解的方程就是在 x_i 处的 Newton 方程.

注意到 Brown 方法中 $L_j(t_{j+1}-\xi_{j+1},\cdots,t_n-\xi_n)(j=1,2,\cdots,k-1)$,则 $t_i-\xi_i$ 是线性的.如果不是把它们代入 $f_k(t)$,而是直接代入

$$f_k(x_i)+f_k'(x_i)^{\mathrm{T}}(t-x_i)=0 \tag{C-1-11}$$

由于 $t_j-\xi_j=L_j(t_{j+1}-\xi_{j+1},\cdots,t_n-\xi_n)(j=1,\cdots,k-1)$,则 $t-x_i$ 总可以写成如下的形式

$$t-x_i=(u_k,\cdots,u_n)\begin{pmatrix}t_k-\xi_k\\ \vdots\\ t_n-\xi_n\end{pmatrix}+q \tag{C-1-12}$$

于是

$$f_k'(x_i)^{\mathrm{T}}(t-x_i)=(f_k'(x_i)^{\mathrm{T}}u_k,\cdots,f_k'(x_i)^{\mathrm{T}}u_n)\begin{pmatrix}t_k-\xi_k\\ \vdots\\ t_n-\xi_n\end{pmatrix}+f_k'(x_i)^{\mathrm{T}}q$$

如用差商代替导数,则不必计算 $f_k'(x_i)$,而利用在 $u_k,\cdots,u_n,q$ 上差商,即

$$f_k'(x_i)^{\mathrm{T}}u_l\approx\frac{f_k(x_i+hu_l)-f_k(x_i)}{h}\quad(l=k,\cdots,n)$$

同样解得

$$t_k-\xi_k=L_k(t_{k+1}-\xi_{k+1},\cdots,t_n-\xi_n)$$

相对于通常的离散 Newton 法，其函数值的计算量也得以减少．利用这种离散的方法求解方程(C-1-2)式，其函数值的计算量约为通常的离散 Newton 法的一半．这样，得到一类新形式的离散 Newton 法，它不同于 Brown 方法．下面将由零空间生成导出这一类算法，因此称这类算法为零空间上的离散 Newton 法．

二、Newton 算法

设

$$f_j'(x_i)=a_j,\ -f_j(x_i)=b_j \quad (j=1,2,\cdots,n) \tag{C-2-1}$$

则(C-1-2)式可以写成如下的线性方程组

$$a_j^{\mathrm{T}}(y-x_i)=b_j \quad (j=1,2,\cdots,n) \tag{C-2-2}$$

利用线性方程组的解的结构求解(C-2-2)式．由于

$$a_j^{\mathrm{T}}(y-x_i)=b_j \quad (j=1,2,\cdots,k-1) \tag{C-2-3}$$

的解可以表示成解集流形 $S_k=\{y_k+U_kz,\ z\in R^{n-k+1}\}$，这里 y_k 是(C-2-3)式的一个特解，U_k 是$n\times(n-k+1)$矩阵，而 $R(U_k)$是$(a_1,\cdots,a_{k-1})^{\mathrm{T}}$ 的零空间．下面将通过实用可行的方法构造零空间，并求解 Newton 方程(C-1-2)或(C-1-3)．当 $F'(x_i)$不是稀疏矩阵时，可以通过在零空间上离散来提高算法的效能．

设

$$y_1=x_i,\quad U_1=I_n \tag{C-2-4}$$

若 S_k 是(C-2-3)式的解集，显然

$$a_j^{\mathrm{T}}U_k=0 \quad (j=1,2,\cdots,k-1) \tag{C-2-5}$$

且秩 $\mathrm{rank}(U_k)=n-k+1$ 为求解

$$a_j^{\mathrm{T}}(y-y_1)=b_j \quad (j=1,2,\cdots,k) \tag{C-2-6}$$

使用如下的方法．

算法 1

(1)计算

$$c_k=U_k^{\mathrm{T}}a_k \tag{C-2-7}$$

$$c_k=(a_k,a_1,\cdots,a_n)^{\mathrm{T}} \tag{C-2-8}$$

(2)构造

$$p_k=U_kz_k \tag{C-2-9}$$

这里 z_k 的选取使得$c_k^{\mathrm{T}}z_k\neq 0$．

(3)解关于 λ 的方程

$$a_k^{\mathrm{T}}(y_k+\lambda p_k-y_1)=b_k$$

得到

$$\lambda_k=\frac{b_k-\Delta_k}{a_k^{\mathrm{T}}p_k}=\frac{b_k-\Delta_k}{c_k^{\mathrm{T}}z_k} \tag{C-2-10}$$

这里 $\Delta_k=a_k^{\mathrm{T}}(y_k-y_1)$．

(4)令

$$y_{k+1}=y_k+\lambda_kp_k \tag{C-2-11}$$

(5)校正

$$U_{k+1}=U_kV_k \tag{C-2-12}$$

这里 V_k 是$(n-k+1)\times(n-k)$列满秩矩阵,且满足

$$V_k^{\mathrm T}c_k=0 \tag{C-2-13}$$

由(C-2-5)、(C-2-7)、(C-2-13)式有

$$a_j^{\mathrm T}U_{k+1}=a_j^{\mathrm T}U_kV_k=0\quad(j=1,2,\cdots,k) \tag{C-2-14}$$

由于 V_k 列满秩,则 U_{k+1}也列满秩.这样得到(C-2-6)式的解集

$$S_{k+1}=\{y_{k+1}+U_{k+1}z,z\in R^{n-k}\} \tag{C-2-15}$$

有各种方法选取 V_k,这里使用与列主元 Gauss 消去法或 QR 分解相对应的取法,即通过对 c_k 的 Gauss 变换或 Householder 变换来选取 V_k.

如果对 c_k 作 Gauss 变换,则应先选主元,交换 U_k 中的列,使得 c_k 中的第一个分量α_k 为

$$|\alpha_k|=\max\{|\alpha_i|,i=k,\cdots,n\} \tag{C-2-16}$$

设 $m=\left(0,\frac{\alpha_{k+1}}{\alpha_k},\cdots,\frac{\alpha_n}{\alpha_k}\right)^{\mathrm T}=(0,m_{k+1},\cdots,m_n)^{\mathrm T}$,则 Gauss 初等下三角变换

$$V^{\mathrm T}=I-me_1^{\mathrm T}$$

使得 $V^{\mathrm T}c_k=\alpha_ke_1$,划去 V 的第一列就可得到 V_k:

$$V_k=\begin{pmatrix}-m_{k+1} & \cdots & -m_n\\ 1 & & \\ & \ddots & \\ & & 1\end{pmatrix} \tag{C-2-17}$$

这时,U_k 总可以表示成

$$U_k=\begin{pmatrix}s_{1k} & \cdots & s_{1n}\\ \vdots & & \vdots\\ s_{k-1k} & \cdots & s_{k-1n}\\ & I_{n-k+1} & \end{pmatrix} \tag{C-2-18}$$

当选主元时,需交换 U_k 的列.为了使 I_{n-k+1}形式不变,仅需交换自变量 x 的与之相对应的两个分量之间的次序.

利用(C-2-18)式不难得到,求解(C-2-1)式需$\frac{1}{3}n^3$ 次乘法运算,这与列主元 Gauss 消去法相同;但其存储量仅为$\frac{n^2}{4}$,而列主元 Gauss 消去法的存储量为 n^2.

如果对 c_k 作 Householder 变换,取

$$V=I-\pi^{-1}uu^{\mathrm T}$$

这里 $u=c_k+\sigma e_1,\sigma=\mathrm{sign}(\alpha_k)\|c_k\|,\pi=\frac{1}{2}\|u\|^2$.划去 V 的第一列可得 V_k.事实上,这种方法相当于对$(a_1,a_2,\cdots,a_n)$的 QR 分解,但不存储 R,故仅需存储$\frac{1}{2}n^2$ 个元素.

如果使用 $f_j(x_i)$的差商表示 a_j,由算法 1 所得到的结果与通常的离散 Newton 法完全相同.

如果 $F'(x_i)$ 不是稀疏矩阵,可以通过下面的方法在零空间上进行离散,从而使函数值的计算量得以减少.

在算法 1 中,使用差商代替导数来计算 $U_k^{\mathrm{T}}a_k$. 设 $U_k=(u_k,u_1,\cdots,u_n)$,则

$$U_k^{\mathrm{T}}a_k=(u_k^{\mathrm{T}}f_k'(y_1),\cdots,u_n^{\mathrm{T}}f_k'(y_1))^{\mathrm{T}} \tag{C-2-19}$$

使用方向差商

$$\frac{\Delta f_k(y_1)}{\Delta u_j}=\frac{f_k(y_1+hu_j)-f_k(y_1)}{h}\quad(j=k,\cdots,n) \tag{C-2-20}$$

代替方向导数 $u_j^{\mathrm{T}}f_k'(y_1)(j=k,\cdots,n)$,这样

$$c_k=\left(\frac{\Delta f_k(y_1)}{\Delta u_k},\cdots,\frac{\Delta f_k(y_1)}{\Delta u_n}\right)^{\mathrm{T}} \tag{C-2-21}$$

对 $\Delta_k=a_k^{\mathrm{T}}(y_k-y_1)(k>1)$,设

$$y_k-y_1=s_k,\ u_s=\frac{s_k}{\|s_k\|} \tag{C-2-22}$$

用

$$\Delta_k=\frac{f_k(y_1+hu_s)-f_k(y_1)}{h}\|s_k\| \tag{C-2-23}$$

代替 $f_k'(y_1)^{\mathrm{T}}s_k$.

由(C-2-10)式和 $b_k=-f_k(y_1)$ 有

$$\lambda_k=\frac{-(f_k(y_1)+\Delta_k)}{c_k^{\mathrm{T}}z_k} \tag{C-2-24}$$

为了进一步提高算法的效能,可以选取适当的 r,采用 r 步重新计算一次 Jacobi 阵 $F'(x_i)$. 对算法 1,这等价于多步重新计算一组 $p_j(j=1,2,\cdots,n)$. 设 $x_i=x_i^0$,若 $F'(x_i)$ 不变,得到序列设为 $x_i^l(l=1,2,\cdots,r)$,为了实用起见,设 r 可变,并采用如下的准则

$$\|F(x_i^l)\|\leqslant\beta\|F(x_i^{l-1})\|,\ \beta\in(0,1) \tag{C-2-25}$$

来决定 r 的取值. 当(C-2-25)式成立时,可重新使用前一组 p_j.

综上所述,完整的零空间上的离散 Newton 法如下:

算法 2

(1)置 $y_1=x_i,U_1=I,x_i^l=x_i,l=0,k=1$.

(2)计算方向差商.

$$c_k=(\alpha_k,\alpha_{k+1},\cdots,\alpha_n)^{\mathrm{T}}$$

这里

$$\alpha_j=\frac{f_k(x_i+hu_j)-f_k(x_i)}{h}\quad(j=k,\cdots,n) \tag{C-2-26}$$

$u_k,u_{k+1}\cdots,u_n$ 是 U_k 的列向量.

(3)选取 z_k,使得 $c_k^{\mathrm{T}}z_k\neq0$.

(4)计算 $p_k=U_kz_k$.

(5)对 $k=1$,取 $\Delta_1=0$;对 $k>1$,计算 $s_k=y_k-y_1,u_s=\dfrac{s_k}{\|s_k\|}$,

$$\Delta_k = \frac{f_k(x_i + hu_s) - f_k(x_i)}{h} \| s_k \|$$

$$\lambda_k = \frac{-(f_k(y_1) + \Delta_k)}{c_k^{\mathrm{T}} z_k} \tag{C-2-27}$$

$$y_{k+1} = y_k + \lambda_k p_k$$

(6)若 $k = n$ 则转(9).

(7)构造$(n-k+1)\times(n-k)$列满秩矩阵 V_k，使得 $V_k^{\mathrm{T}} c_k = 0$.

(8)校正

$$U_{k+1} = U_k V_k$$

置 $k = k+1$，转(2).

(9)置 $y_{n+1} = x_i^{l+1}$.

(10)对 $\beta \in (0,1)$，若 $\| F(x_i^{l+1}) \| \leqslant \beta \| F(x_i^l) \|$，则置 $y_1 = x_i^{l+1}$，按(5)中的公式依次计算 $y_2, \cdots, y_{n+1}$，置 $l = l+1$，转(9)；否则置

$$x_{i+1} = \begin{cases} x_i^l, & l = 0 \\ x_i^l, & l \geqslant 1 \end{cases}$$

$i = i+1$，转(1).

算法 2 实际上可以看成求解如下的线性方程：

$$F(x_i) + A(x_i, h)(x - x_i) = 0 \tag{C-2-28}$$

且有

$$\lim_{h \to 0} \| F'(x_i) - A(x_i, h) \| = 0 \tag{C-2-29}$$

由(C-2-29)式可知，当 $h \to 0$ 时，算法 2 就是通常的 Newton 法. 而当 $h \to 0$ 时，Brown 方法求解的方程是(C-1-10)，(C-2-29)式则不成立，从而得不到通常的 Newton 方程，在解点附近只是 Newton 方程的一个近似.

相对于 $F(x)$的模长而言，Newton 方向是一个下降方向，适当地选取步长，算法 2 可以有大范围的收敛性质. 而 Brown 方法显然不具有这一性质.

在非线性方程组中，如果其中部分方程是线性方程，不妨设 $f_1(x), \cdots, f_{k-1}(x)$是线性的，这时用零空间法求解 Newton 方程，仅需在流形 $S_k = \{ y_k + U_k z, z \in R^{n-k+1} \}$ 上求解非线性方程组 $f_j(x) = 0 (j = k, \cdots, n)$. 相对通常的 Newton 法，其效能也得到了提高，如下例：

$$f_j(x) = -(n+1) + x_j + \sum_{i=1}^{n} x_i,\ j = 1, \cdots, n-1,\ f_n(x) = -1 + \prod_{i=1}^{n} x_i \tag{C-2-30}$$

$n = 50$，初值取为$\left(\frac{1}{2} \quad \cdots \quad \frac{1}{2} \right)^{\mathrm{T}}$，用双精度对前 $n-1$ 个线性方程一次迭代，产生了一维流形 $S_n = \{ y_n + U_n z, z \in R \}$. 在一维流形 S_n 上求解 $f_n(x) = 0$，仅 9 次迭代得到解 x^*，其精度为 $\| F(x^*) \| \approx 10^{-13}$.

三、数值检验

（略）

四、收敛性讨论

（略）

参考文献

（略）

（原载《计算数学》第二十卷第一期，1998）

参 考 文 献

[1] 成平,陈希孺,陈桂景,吴传义.参数估计.上海:上海科学技术出版社,1985
[2] 王松桂.线性模型的理论及其应用.合肥:安徽教育出版社,1987
[3] 陈希孺,王松桂.近代回归分析.合肥:安徽教育出版社,1987
[4] 武汉测绘科技大学测量平差教研室.测量平差基础.北京:测绘出版社,1996
[5] 杨元喜.抗差估计理论及其应用.北京:八一出版社,1993
[6] 罗永光,王海云.稳健信号处理概论.北京:国防科技大学出版社,1987
[7] 周江文,黄幼才,杨元喜,欧吉坤.抗差最小二乘法.武汉:华中理工大学出版社,1997
[8] 王新洲.基于信息扩散原理的估计理论、方法及其抗差性.武汉测绘科技大学学报,1999(3):240~244
[9] 黄崇福.模糊信息优化处理技术在中国发展的18年.(2002,待发表)
[10] 黄崇福,王家鼎.模糊信息优化处理技术及其应用.北京:北京航空航天大学出版社,1995
[11] 游扬声.信息扩散原理及其在测量数据处理中的应用.武汉大学(学位论文),2001
[12] 王新洲,游扬声.论信息扩散估计的窗宽.测绘科学,2001(1):16~19
[13] 游扬声,王新洲.适用于高斯-马尔可夫模型的信息扩散估计及其性能.测绘科学,2001(2):15~19
[14] 王新洲.用信息扩散估计进行水准网平差.武汉测绘科技大学学报,2000(5):405~407
[15] Wang Xinzhou. Adjustment of Leveling Network by Information Spread Estimation. Geo-Spatial Information Science, 2001(1), pp. 1-4
[16] 王新洲.在无偏估计类中改进最小二乘估计的方法.武汉测绘科技大学学报,1995(1):46~50
[17] 王新洲,刘丁酉,张前勇,黄海兰.谱修正迭代法及其在测量数据处理中的应用.黑龙江工程学院学报,2001(2)
[18] 刘丁酉.矩阵分析.武汉:武汉测绘科技大学出版社,1998
[19] 於宗俦,鲁林成.测量平差基础(增订本).北京:测绘出版社,1983
[20] 张金槐,唐雪梅.BAYES方法.北京:国防科技大学出版社,1989
[21] 王新洲.广义平差的概括模型.武汉测绘科技大学学报,2000(3):257~260
[22] 焦李成.神经网络系统理论.西安:西安电子科技出版社,1990
[23] 国家自然科学基金委员会.大地测量学.北京:科学出版社,1994
[24] 韦博成.近代非线性回归分析.南京:东南大学,1989
[25] P. J. G. Teunissen. The geometry of geodetic inverse linear mapping and non-linear

adjustment. Netherlands Geodetic Commission, Publ. On Geodesy, New Series. Vol. 8 (1). Delft, 1985

[26] P. J. G. Teunissen and E. H. Knickmeyer. Non-linearity and least squares, CISM Journal ACSGC. Vol. 42(4), pp. 383-390, 1988

[27] P. J. G. Teunissen. First and Second moments of non-linear least-squares estimation. Bull Geod. 13, pp. 253-262, 1989

[28] P. J. G. Teunissen. Nonlinear least squares. Manuscript Geodaetica, Vol. 15, pp. 137-150, 1990

[29] Blaha. G. Non-iterative approach to nonlinear least-squares adjustment. Manuscript Geodaetica. 1994(19), pp. 199-212

[30] Lohse. P. Parameter Estimation in nonlinear Models(in German). Ph. D. dissertation, Stuttgart University. Det. of Geodetic Science. Stuttgart Federal Republic of Germany, 1993

[31] Athanasios Dermanis and Fernando Sanso. A study of non-linear estimation, IUGG 交流论文, 1995

[32] 徐培亮.非线性函数的协方差传播公式.武汉测绘科技大学学报,1986(2):92～99

[33] 刘大杰,黄加纳.非线性最小二乘平差的迭代解法.武测科技,1987(4)

[34] 周世健.广义方差——协方差传播律.大地测量学的发展.北京.测绘出版社,1996.

[35] 刘国林,陶华学.非线性观测值函数的协方差和协因数传播及其权例数.测绘工程. 1997(2):8～16

[36] Liu Guolin, Guo Tinyun, Tao Huaxue. Two Kinds of Explicit Methods to Nonlinear Adjustments of Free-Networks with Rank Deficiency. Geomatics Research Australasia No. 73, pp. 25-32, 2000

[37] 胡圣武.非线性模型估计及其在 GIS 中的应用.武汉测绘科技大学(学位论文),1997

[38] 胡圣武,陶本藻.非线性模型的误差传播及其在 GIS 中的应用.武汉测绘科技大学学报,1997(2):129～131

[39] 王新洲.非线性模型线性近似的容许曲率.武汉测绘科技大学学报,1997(2):119～121

[40] 王新洲.非线性模型参数估计的直接解法.武汉测绘科技大学学报,1999(1):64～67

[41] 王新洲.非线性模型能否线性化的实用判据.武汉测绘科技大学学报,1999(2):145～148

[42] 王新洲.基于信息扩散原理的估计理论、方法及其抗差性.武汉测绘科技大学学报,1999(3):240～244

[43] 王新洲.非线性模型平差中单位权方差的估计.武汉测绘科技大学学报,2000(4):358～361

[44] 李朝奎,徐望国,曾卓乔,黄力民.非线性函数空间平差方程的解法及其特征.测绘学院学报,Vol. 18, No. 1, 2001:8～11

[45] 王新洲.测量平差.北京:水利水电出版社,1991

[46] 李庆海,陶本藻.概率统计原理和在测量中的应用.北京:测绘出版社,1982

[47] 黄维杉.近代平差理论及其应用.北京:解放军出版社,1992
[48] 邓乃扬等.无约束最优化计算方法.北京:科学出版社,1982
[49] 王德人.非线性方程组解法与最优化方法.北京:人民教育出版社,1979
[50] Levenberg.K. A method for the solution of Certain nonlinear problems in least squares. Quart.Appl.Math.2,1944,pp.164-168
[51] Marquardt, D. W., An algorithm for least squares estimation of nonlinear parameters. SIAM J.,11,1963,pp.431-441
[52] Fletcher.R.A modified Marquardt subroutine for nonlinear least squares. AERE, Marwell. Report, No. R-6799,1971
[53] 熊西文,施吉林,谢如微.数值计算.武汉:华中理工大学出版社,1986
[54] 高士纯,于正林.测量平差基础习题集.北京:测绘出版社,1983
[55] 陈宝林.最优化理论与算法.北京:清华大学出版社,1989
[56] 林洪桦.动态测试数据处理.北京:北京理工大学出版社,1995
[57] 李士勇.模糊控制·神经控制和智能控制.哈尔滨:哈尔滨工业大学出版社,1998
[58] 陈国良,王熙法,庄镇泉等.遗传算法及其应用.北京:人民邮电出版社,1996
[59] 李德仁.误差处理和可靠性理论.北京:测绘出版社,1988
[60] J.Saleh.Robust estimation based on energy minimization principles.Journal of Geodesy,(2000)74
[61] 周江文.经典误差理论与抗差估计.测绘学报,Vol.18,No.2,1989
[62] 林路.非线性模型中拟似然估计的若干性质.应用数学学报,1999(2):307～310
[63] 陈志,高旅端,邓乃扬.解非线性方程组的一类离散的 Newton 算法.计算数学,1998(1):57～68
[64] 张尧庭,陈汉峰.贝叶斯统计推断.北京:科学出版社,1991
[65] 陈希孺.数理统计中的两个学派——频率学派和 Bayes 学派.数理统计与应用概率,1990(4)
[66] 成平.对贝叶斯统计的几点看法.数理统计与应用概率,1990(4)
[67] 王新洲.稳健二次估计理论及其在 GPS 随机模型估计中的应用.武汉测绘科技大学(学位论文),1994
[68] 张双林,沙秋英,程美玉.回归函数非线性小波估计的一致强相合性.应用概率统计,1999(4):375～380
[69] Hall, P. and Patil, P. On the choice of smoothing parameter, thresholding and trunction in nonparametric regression by nonlinear methods, Research Report, SMS-72-93, School of Mathematical Science. The Australian National University, 1993
[70] 朱仲义,韦博成.半参数非线性回归模型渐近推断的几何.应用数学学报,2001(1):87～99
[71] 吴俊昶,刘大杰,于正林.控制网测量平差.北京:测绘出版社,1998
[72] 李德仁,金为铣,尤兼善,朱宜萱.基础摄影测量学.北京:测绘出版社,1995
[73] 朱肇光,孙护,崔炳光.摄影测量学.北京:测绘出版社,1995
[74] 周忠谟,易杰军,周琪.GPS 卫星测量原理与应用(第二版).北京:测绘出版社,1997

[75] Bates, D. M. and Watts, D. G. Relative curvature measure of nonlinearity. J. R. Statist. Soc. B 42, pp. 1-25, 1980

[76] Tsai, C. L. Contributions to the design and analysis of nonlinear models. Ph. D. Thesis. Univ. of Minnesota. 1983

[77] 韦博成.非线性回归模型 LS 估计量的二阶矩.高校应用数学学报.1986(1):279～285

[78] 徐钟济.蒙特卡罗方法.上海:上海科学技术出版社,1985

[General Information]
□□=□□□□□□□□□□□□□□
□□=□□□□
□□=137
SS□=10943115
DX□=
□□□□=2002□06□□1□
□□□=□□□□□□□